

PAPER

Ecosystem-Driven Design Thinking: A Milestone-Based Progression and Attrition Analysis of Innovation Competence

Dely Lazo-Barreda¹  ,
Elizabeth-Susan Mamani-
Machaca² , Karina
Rosas-Paredes¹ ,
Midwar Ancco¹ ,
Carla Cuya-Zevallos¹ 

¹Universidad Católica
de Santa María, Arequipa,
Peru

²Universidad San Ignacio
de Loyola, Lima, Peru

dlazo@ucsm.edu.pe

ABSTRACT

Universities increasingly adopt SDG-anchored challenge-based learning (CBL) to develop innovation competencies, yet evaluations often rely on self-reports that cannot distinguish enthusiasm from demonstrated competence. This study evaluates the Innovation Challenge, an eight-week ecosystem-driven CBL (Ed-CBL) program integrated with Design Thinking (DT), and is reported using the TIDieR checklist to enhance replicability. We employed a convergent mixed-methods design: primary evidence comprised team-level performance across four mandatory deliverables and a summative pitch evaluated by an expert jury; secondary evidence assessed individual skill gains using the Youth Innovation Skills Measurement Tool (YISMT) in a non-equivalent groups pre- and post-test design. A milestone-based attrition analysis reveals a rigorous selection filter: only 55.5% (15/27) of verified teams overcame ecosystem constraints to complete the full cycle. Statistical analysis of skill gains (Welch's t-test with Holm-Bonferroni correction) identified an initial differential gain solely in Creativity. However, a sensitivity analysis adjusting for significant baseline imbalances rendered this effect non-significant ($p = .267$); other dimensions also showed no significant changes. Consequently, we conclude that while Ed-CBL generates verifiable design outputs and technical performance, it does not yield robust short-term gains in self-reported innovation skills, highlighting the necessity of prioritizing performance evidence over self-efficacy in short-term programs.

KEYWORDS

challenge-based learning (CBL), design thinking, engineering education, mixed methods, innovation competency, sustainable development goals

Lazo-Barreda, D., Mamani-Machaca, E.-S., Rosas-Paredes, K., Ancco, M., Cuya-Zevallos, C. (2026). Ecosystem-Driven Design Thinking: A Milestone-Based Progression and Attrition Analysis of Innovation Competence. *International Journal of Engineering Pedagogy (iJEP)*, 16(4), pp. 69–89. <https://doi.org/10.3991/ijep.v16i4.61279>

Article submitted 2026-02-25. Revision uploaded 2026-05-16. Final acceptance 2026-05-16.

© 2026 by the authors of this article. Published under CC-BY.

1 INTRODUCTION

Engineering programs face increasing pressure to prepare graduates capable of addressing socio-technical challenges aligned with the Sustainable Development Goals (SDGs). In these scenarios, solutions must satisfy technical feasibility under strict regulatory and economic constraints [1], [2]. Consequently, pedagogies such as challenge-based learning (CBL) have gained traction to move education from classroom simulations to authentic problem-solving [3]. However, critics argue that many CBL formats remain “protected,” lacking the uncertainty and boundary conditions that characterize actual professional practice [1], [4].

To close this gap, the ecosystem-driven CBL (Ed-CBL) model proposes integrating external actors not merely as observers but as sources of active constraints [5]. In contexts of water scarcity, for instance, the viability of an engineering solution depends not only on physics but also on governance, where infrastructure and regulatory frameworks define the limits of possibility [6], [7]. Despite its promise, current literature often uses the term “ecosystem” descriptively, without fully operationalizing how “friction with reality,” defined here as direct stakeholder feedback and the rejection of unfeasible ideas, acts as an instructional mechanism to filter and refine technical competence [8]. A compounding methodological barrier is the field’s overreliance on self-reported outcomes—such as satisfaction or self-efficacy—to infer innovation skill development [9]. While informative for motivational constructs, these self-reports are vulnerable to social desirability biases and novelty effects. Such biases obscure the distinction between transient enthusiasm and actual procedural competence, often inflating perceived abilities without correlating to objective performance [10]. For engineering education, this limitation is decisive: the credibility of “innovation competency” must rest on verifiable evidence that students can generate feasible artifacts under severe constraints, rather than on self-reported perceptions of improvement [4], [11]. This shift necessitates authentic assessment and evidence-centered design (ECD) [12], [13].

Furthermore, cumulative knowledge is hindered by the incomplete reporting of intervention details. CBL programs are complex, multi-component interventions whose efficacy depends on implementation fidelity and the alignment of curricular elements. When studies omit essential parameters, such as teacher roles, grouping strategies, or specific materials, replication becomes impossible, turning the pedagogy into a “black box” where outcomes may be driven by unreported confounds rather than instructional design [8]. To ensure reproducibility, this study adheres to the TIDieR (Template for Intervention Description and Replication) framework, explicitly documenting the active components to address the lack of consensus on CBL implementation [8], [14], [15].

To bridge these gaps, this study evaluates the Innovation Challenge 2025, an Ed-CBL intervention anchored in a “wicked problem” defined by the water industry. Using a convergent mixed-methods design, the study pursues three specific objectives: (i) to document implementation fidelity and milestone-based attrition through sequential deliverables; (ii) to characterize the quality of performance artifacts as validated by expert juries; and (iii) to triangulate these results with psychometric gains, measured by the Youth Innovation Skills Measurement Tool (YISMT) within a non-equivalent groups pre- and post-test comparison. This approach shifts the evaluation focus from subjective satisfaction to verifiable evidence of performance under real-world constraints.

2 THEORETICAL FRAMEWORK

2.1 Ecosystem-driven CBL: From academic simulation to “wicked problems”

Unlike traditional project-based learning (PBL), where problems are typically pre-defined and contained within the classroom, CBL grounds itself in exposing students to “wicked problems” anchored in the SDGs and situated within authentic ecosystems [8], [16]. The Ed-CBL model posits authentic challenges, defined through interaction with water sector stakeholders (e.g., “Optimizing drinking water in scarcity zones”), activate learning conditions that exceed academic simulation and situate student innovation under strict technical, operational, and regulatory constraints [4].

Literature defines CBL effectiveness not merely by task resolution but by students’ ability to navigate uncertainty and constraints imposed by external actors (industry, regulators, and NGOs) [4], [8]. Consequently, the magnitude of the learning effect depends on the “dose” of ecosystem exposure, defined here as the frequency of stakeholder sessions and regulatory feedback. For example, interaction with stakeholders such as SEDAPAL or SUNASS is not accessory; it constitutes a critical validation mechanism where feedback provides real “boundary conditions” (technical and economic feasibility) that cannot be simulated in a closed environment [17]. This pedagogical architecture transforms students from passive recipients into epistemic agents who co-construct solutions within the ecosystem, aligning their competencies with global sustainability demands [4], [8], [18]. Learning emerges through iterative cycles of research and validation; thus, increased exposure to ecosystem constraints (“constructive friction”) correlates positively with final performance quality. These constraints act as a selective filter, ensuring only viable proposals progress through the validation gates [8], [16].

This human-centric validation is critical. Recent evidence from Western higher education demonstrates that delegating the evaluation of complex, open-ended coursework to generative AI tools carries severe invalidity risks [13], [19]. Empirical studies using Bland-Altman analyses reveal that AI exhibits a systematic proportional bias. This bias artificially inflates scores for weaker submissions and compresses variance, failing to reliably discriminate genuine competency across the performance spectrum [19].

2.2 The procedural mechanism: Design thinking (DT) and the double diamond

Within the macro-context of CBL, DT operates as the procedural micro-mechanism structuring the student’s response to ambiguity. Far from being treated solely as a creativity technique, DT is operationalized here as a complex problem-solving competency that generates tangible evidence of learning [8], [15]. Each phase of the cycle (Empathize, Define, Ideate, Prototype, Test) functions as an epistemic value chain, producing observable artifacts (e.g., empathy maps, value proposition canvases, low- and high-fidelity prototypes). These artifacts are not mere administrative deliverables; they represent the external materialization of cognitive processes and serve as “boundary objects” for meaning negotiation between students and mentors, transforming abstract cognition into assessable evidence [8], [16].

The model adopts the Double Diamond architecture, divided into four iterative phases aligned with the intervention milestones (E1–E4) [20], [21] (see Figure 1).

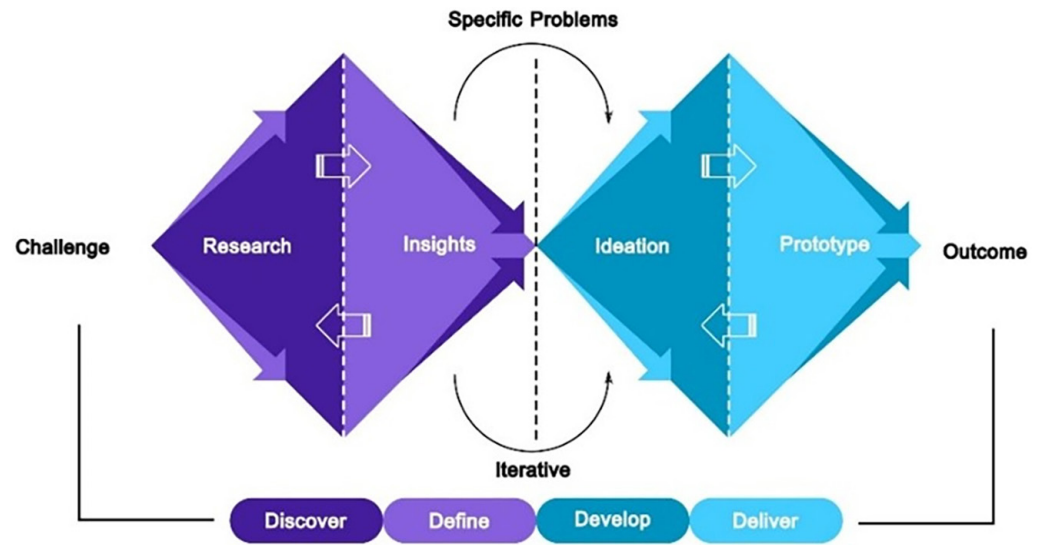


Fig. 1. Double diamond model adapted for the innovation challenge

Specifically, the first diamond (Problem Space) forces an initial divergence through field research and a convergence toward a falsifiable hypothesis (Experiment Canvas), with the critical objective of avoiding “premature solution bias” via problem reframing [15], [21]. The second diamond (Solution Space) spans from divergent ideation (application of SCAMPER) to technical convergence, where tools such as the MoSCoW matrix and the MVP Canvas act as viability filters, ruling out features unsustainable under the regulatory constraints of the water ecosystem [15].

The framework posits that learning and innovation emerge from the tension inherent in these divergence-convergence cycles [20]. Consequently, team attrition is considered not a failure of instructional design but evidence of the filtering mechanism’s fidelity: only proposals that overcome these constraints in the second diamond (technical validation) transition from ideation to operational feasibility. This iterative process explains why skill gains are expected to be more pronounced in procedural competence than in stable attitudinal traits, given the intensive focus on artifact iteration [8], [16].

2.3 Assessment hierarchy: Primacy of performance evidence

To mitigate biases inherent in perception-based studies (e.g., ceiling effects and social desirability), this framework establishes a strict epistemological hierarchy: performance evidence is primary, while self-reports are supplemental. Grounded in Authentic Assessment principles, the model posits that innovation competence must be inferred from the technical quality of artifacts produced in realistic contexts, rather than self-confidence alone. Consequently, evaluation focuses on “what the team produces” under ecosystem constraints, utilizing rubrics aligned with professional standards to quantify demonstrated competence [11], [15].

Methodological triangulation strengthens the study’s internal validity by contrasting these objective outputs against subjective perceptions [22]. While self-report instruments are useful for measuring motivational variables [1],

they remain insufficient to capture the complexity of applied engineering competence [15], [23]. Furthermore, team attrition is incorporated as a metric of implementation fidelity: the ability of teams to clear technical “quality gates” constitutes objective evidence of rigor, validating the program architecture beyond mere participant satisfaction.

2.4 Research hypotheses

Based on the Ed-CBL architecture and the assessment hierarchy described above, this study posits two directional hypotheses:

- H1: Teams exposed to the Ed-CBL program will successfully internalize real-world constraints, producing innovation artifacts that meet the rigorous technical and operational thresholds validated by an external expert jury.
- H2: Program participation will be associated with positive gains in self-reported innovation skills, with the largest short-term effects expected in malleable procedural competencies, contrasting with smaller or negligible changes in attitudinal traits over the eight-week period.

3 MATERIALS AND METHODS

3.1 Study design and reporting standards

This study employed a convergent mixed-methods design, simultaneously collecting and analyzing group performance data (innovation artifacts) and individual self-perception metrics (pre and post-test surveys) [24]. To ensure technical reproducibility and methodological transparency, addressing the “black box” criticism common in educational interventions, the intervention reporting strictly adheres to the TIDieR (Template for Intervention Description and Replication) checklist (refer to Table A1 in Appendix A). Additionally, the quasi-experimental component for skills assessment follows TREND standards for non-randomized evaluations [25]. This rigorous reporting framework differentiates this study from anecdotal reports, providing a solid foundation for theoretical generalization.

3.2 Participants and context

The intervention was conducted at a private university located in southern Peru. The study context spanned an intensive cycle from kick-off to final pitch (March 25–June 20, 2025).

To control selection bias and ensure cognitive diversity, a “mandatory multidisciplinary” criterion was applied, requiring each team to include members from at least three distinct academic disciplines. After this validation, 27 teams formally began the intervention (Appendix A, Table A1). For the analysis of latent competencies (YISMT), participants with incomplete data at T2 were excluded, resulting in a final analytical sample of $N = 91$ undergraduate students ($n_{\text{experimental}} = 61$; $n_{\text{control}} = 30$).

To construct the non-equivalent control group, participants were drawn exclusively from the pool of volunteers who registered for the Ed-CBL program but were excluded due to cohort capacity constraints. This approach controlled for self-selection bias regarding initial interest and motivation.

Ethical considerations. Participation was voluntary, and written informed consent was obtained for the use of anonymized data, in accordance with COPE guidelines and the Declaration of Helsinki. The program was approved by the institution under Resolution N°. 31376-R-2025. To mitigate coercion risks, a key concern in educational research, academic participation was strictly dissociated from research participation; data collection had no impact on course grades, ensuring that refusal to participate carried no academic penalty [26].

3.3 The intervention: Ed-CBL architecture

The “Innovation Challenge 2025” was designed to operationalize the Ed-CBL model described in the theoretical framework. The pedagogical objective was to subject multidisciplinary teams to an intensive 8-week cycle of problem-solving under uncertainty. Unlike traditional CBL, this design explicitly introduced “friction with reality” as a mediating variable through three standardized operational layers, documented according to the TIDieR protocol [27] (refer to Appendix A, Table A1):

Ecosystem layer (inputs – boundary constraints). The challenge was not an academic simulation but a “wicked problem” defined by the water industry (SEDAPAL) regarding the “Optimization of drinking water in scarcity zones” [4], [28]. This layer imposed exogenous “boundary conditions” (SUNASS regulations, infrastructure limitations, and operational costs), forcing teams to discard theoretically desirable but unfeasible solutions during validation sessions.

Instructional layer (process – the dose). Following the Double Diamond architecture (see Figure 2), the intervention was managed through four Synchronous Workshops (4 hours each) and a sequence of four mandatory deliverables (Gatekeeping milestones) that functioned as fidelity control points [8]. In the Detection phase (E1, Weeks 1–2), teams constructed Problem Trees and conducted field validation to avoid early solution bias [29]. Subsequently, the Ideation phase (E2, Weeks 3–4) required the mandatory application of SCAMPER and Customer Journey Maps to enforce-controlled divergence [30]. The critical Feasibility phase (E3, Weeks 5–6) demanded technical convergence using the MoSCoW matrix (Must/Should/Could/Won’t), where teams filtered proposals based on ecosystem constraints [31]. Finally, the Prototyping phase (E4, Weeks 7–8) focused on developing the Minimum Viable Product (MVP Canvas) and defending it via an impact pitch [32]. The specific evaluation criteria for each milestone are detailed in the process rubrics (refer to Appendix B, Table B1).

Support layer (scaffolding). External experts and trained teachers conducted four specialized mentoring sessions (1 hour per milestone per team). These sessions were not merely consultative; they acted as “quality gates” designed to reduce the variance between academic proposals and regulatory feasibility prior to the final summative evaluation.

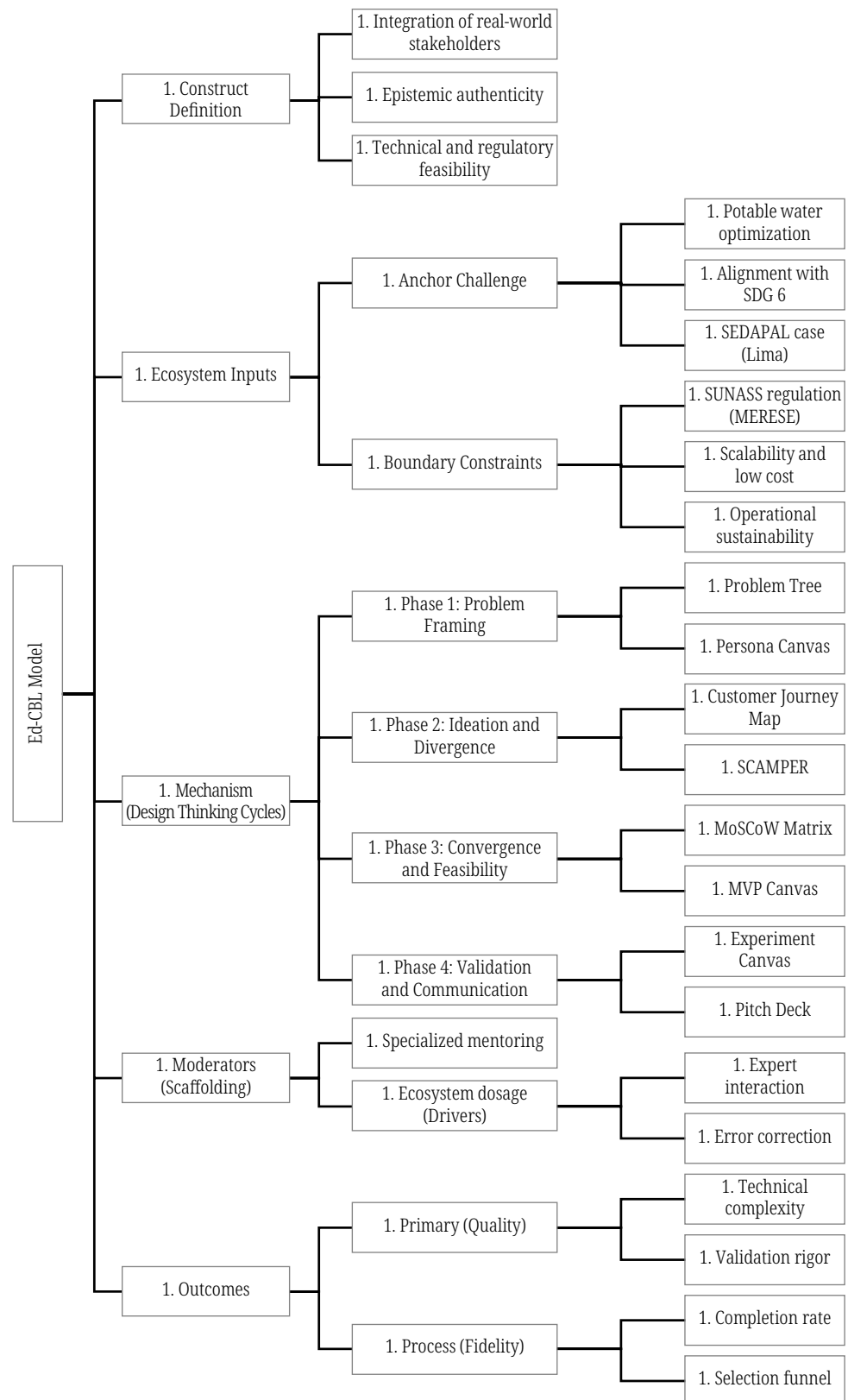


Fig. 2. Conceptual framework of the intervention mechanisms and expected outcomes

Notes: The mechanism follows a Double Diamond progression through four phases. “Ecosystem Dosage” is defined as the intensity of exposure to ecosystem constraints, mediated by expert interaction and immediate error correction cycles.

3.4 Data collection and instruments

Primary outcome: Performance (artifact quality). Innovation competency was evaluated via the technical quality of final prototypes using a standardized analytic rubric with six weighted dimensions: Innovation, Problem Identification, Viability, Impact, Scalability, and Pitch Presentation. To ensure objectivity and mitigate conflicts of interest, the summative assessment was conducted by an independent external jury composed of experts from the National Water Authority (ANA), SEDAPAR, and business incubators. The evaluation followed a “single-blind” design (evaluators had no prior mentoring roles with assigned teams) to ensure impartiality in the scoring (0–20 scale) [33], [34]. The complete summative evaluation rubric is detailed in Appendix B (refer to Table B2).

To establish the psychometric rigor of the summative assessment, the inter-rater reliability of the external panel was evaluated a priori using the Intraclass Correlation Coefficient (ICC), based on a two-way mixed, average-measures model [35], [36].

This primary outcome was assessed exclusively within the experimental cohort ($n = 61$; 15 teams). As the control group curriculum did not include the Double Diamond convergence cycles and the prototype development phase, no artifact evaluation was conducted for this cohort.

Secondary outcome: Individual innovation skills. As a triangulation measure, self-perception of innovation skills was assessed using the YISMT [37] at two time-points (T1: start, T2: end). Internal reliability was verified for each measurement wave, reporting consistent coefficients (Cronbach’s $\alpha > 0.90$), confirming psychometric robustness. Detailed reliability tables and validity checks are available in the [Supplementary Material \(refer to Section 1\)](#).

Process fidelity. The completion rate was recorded as an indicator of implementation fidelity. Monitoring the “participant funnel” allowed us to document model rigor, identifying that 44.5% of verified teams did not overcome the ecosystem constraints. This attrition data is critical for the milestone-based attrition analysis and validates the “filtering” function of the Ed-CBL architecture (refer to Appendix B, Table B1).

3.5 Statistical analysis

Data processing and statistical inference were performed using Python (version 3.12.12). Prior to hypothesis testing, data distributions and assumptions were evaluated; normality was assessed using the Shapiro-Wilk test, and homoscedasticity was examined via Levene’s test.

Baseline comparability and cluster-level variance. To assess baseline equivalence (T1) between the non-equivalent experimental and control groups, the non-parametric Mann-Whitney U test was employed due to the violation of the normality assumption across baseline dimensions. Additionally, because the experimental intervention required participants ($n = 61$) to operate within 15 nested teams, the assumption of independence was formally evaluated. A one-way random-effects ANOVA was used to calculate the team-level Intraclass Correlation Coefficient (ICC1) for the outcome variables. Because ICC1 represents a variance-partitioning

estimate, any negative empirical values derived from the model were interpreted a priori as effectively zero, indicating an absence of meaningful cluster-level variance. This step isolated the variance attributable to group membership, thereby supporting the use of individual-level parametric statistics for the gain scores.

Analysis of individual skill gains. To assess skill gains ($\Delta = T2 - T1$), we employed Welch's t-test for independent samples. This test was selected over the standard Student's t-test due to its methodological robustness in managing heteroscedasticity in unequal sample sizes ($n_{exp} \neq n_{ctrl}$) [38]. The magnitude of the intervention effect was estimated using Glass's Δ rather than Cohen's d ; this estimator provides a more reliable comparator by standardizing the mean difference against the control group's standard deviation, which remained unaffected by the intervention [39]. Finally, to mitigate Type I error inflation derived from multiple comparisons across the six YISMT subscales, the Holm-Bonferroni sequential correction was strictly applied to all p -values [40]. Furthermore, to evaluate the robustness of the findings and account for any potential baseline imbalances inherent to the non-randomized design, a sensitivity analysis was established. Specifically, an ordinary least squares (OLS) regression (Analysis of Covariance, ANCOVA) with heteroscedasticity-robust standard errors (HC3) was planned to adjust post-test scores for baseline covariates, ensuring that any observed differential gains were not statistical artifacts of initial group differences.

4 RESULTS

4.1 Process fidelity and milestone-based attrition analysis

The intervention exhibited a marked implementation gradient consistent with a deliverable-based filter model. Initially, 48 teams were registered, of which 27 (56.2%) met the verification criteria (mandatory multidisciplinary) and formally entered the pipeline.

Protocol adherence functioned as a selective mechanism. Detailed analysis of the "dose received" reveals three performance cohorts among the verified teams ($n = 27$): 15 teams (55.5%) completed the full sequence (Deliverables 1–4 and Final Pitch); 3 teams (11.1%) completed three deliverables; and 9 teams (33.3%) discontinued participation during the early phases. A milestone-based attrition analysis confirms that the highest attrition (33.3%) occurred during the initial technical validation phases, coinciding with the introduction of ecosystem regulatory constraints. This pattern indicates that progress required sustained execution through iterative milestones rather than mere registration.

4.2 Performance evidence: Artifact evolution

Through the four-tiered milestones, teams generated a traceable design chain progressing from problem framing to feasibility logic. The artifacts—ranging from Problem Trees to MVP Canvases—collectively documented iterative divergence-convergence cycles and increasing specificity in design requirements (see Figure 3).

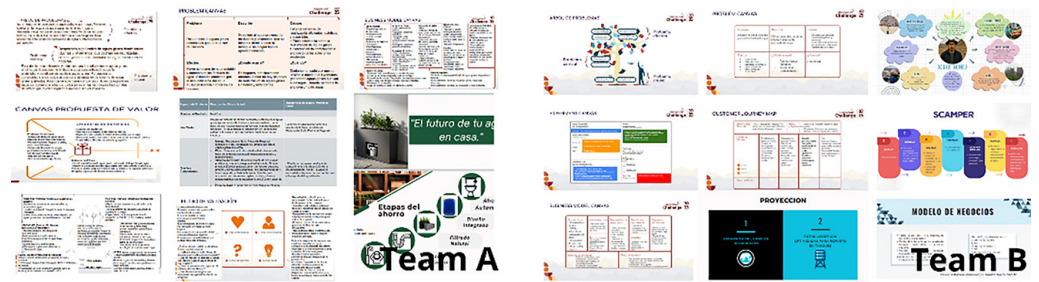


Fig. 3. Progression of artifact fidelity across milestones (E1 → E4)

Content analysis of these deliverables reveals the emergence of two predominant solution families: (a) Digital/IoT Solutions, focusing on monitoring and optimization (e.g., leak sensors, consumption analytics) dependent on field connectivity; and (b) Nature-Based Solutions, characterized by low-cost approaches prioritizing local maintainability (e.g., biological filtration systems).

Regarding technical convergence, the MoSCoW matrices showed a systematic exclusion pattern. Features classified as “Won’t Have” consistently included high-cost imported components or requirements for continuous connectivity difficult to sustain under field constraints. This exclusion pattern indicates a shift in design priorities from abstract creativity to operational feasibility mandated by ecosystem constraints.

4.3 Summative assessment: Expert jury evaluation

The inter-rater reliability analysis of the scores provided by the five-member external expert panel for the 15 finalist teams yielded an ICC of 0.80 (95% CI [0.41, 0.93], $p < .001$).

Summative evaluation scores ranged from 10.5 to 17.5 on the 20-point scale, revealing a distinct performance gradient. As detailed in Table 1, the top-ranked group (Team 5) achieved a consolidated score of 17.5/20, followed by Team 2 (14.9) and Team 13 (14.7). These scores indicate that the top quartile of the cohort successfully met the technical thresholds set by the ecosystem partners (for the complete ranking and disaggregated scores, see [Section 2 in Supplementary Material](#)).

Table 1. Top five performing teams (jury evaluation)

Rank	Team	Solution Focus	Final Score (/20)
#01	Team 5	Biotechnological optimization for sustainable rural water access	17.5
#02	Team 2	Tech-driven innovation for purpose-driven engineering	14.9
#03	Team 13	Secure water access systems for peri-urban communities	14.7
#04	Team 12	Sustainable technological design for social impact in water scarcity	13.8
#05	Team 3	Low-cost solar filtration and rainwater harvesting	13.7

Notes: Final scores represent the consolidated average from the external expert jury (ANA, SEDAPAR, and industry experts). The maximum possible score was 20. Descriptions are based on the final MVP specifications.

4.4 Individual skill gains (YISMT)

Baseline comparability. Baseline equivalence ($T1$) was evaluated between the experimental ($n = 61$) and control ($n = 30$) groups. The analyses revealed no

statistically significant differences between the cohorts regarding the overall YISMT score ($U = 956.0, p = .732$) or in five of its six dimensions: Self-efficacy ($U = 929.0, p = .909$), Energy ($U = 906.0, p = .943$), Risk Propensity ($U = 1066.5, p = .201$), Leadership ($U = 827.0, p = .459$), and Intent to Innovate ($U = 1053.0, p = .241$) (refer to Supplementary Table S4). A statistically significant difference at baseline was observed solely in the Creativity dimension ($U = 1173.0, p = .029$).

Differential gains in innovation competencies. Prior to hypothesis testing, the cluster-level confounding effects derived from team-based nesting in the experimental group were evaluated. The Intraclass Correlation Coefficients (ICC1) computed for the Delta gain scores ranged from -0.106 to 0.044 across all dimensions (all $p > .05$). Negative ICC1 estimates were interpreted as effectively zero, indicating no meaningful between-team variance for the gain scores. Consequently, team membership accounted for a negligible proportion of variance, fully justifying the use of individual-level gain analyses (refer to Supplementary Table S5).

To triangulate performance evidence with perceived skill development, gain scores ($\Delta = T2 - T1$) were compared between the experimental and non-equivalent control groups. After applying the pre-specified decision rule (Welch's t-test) and controlling for family-wise error with the Holm-Bonferroni adjustment, only Creativity showed a significant differential gain. The experimental group improved significantly more than the control group (Mean Difference = 0.471 , 95% CI [$0.15, 0.80$], $p_{Holm} = .031$), exhibiting a large, standardized effect size (Glass's $\Delta = 0.81$). However, given the statistically significant baseline imbalance observed in this specific dimension ($p = .029$), a sensitivity analysis was conducted to evaluate the stability of this finding. Using an ANCOVA model with robust standard errors (HC3) to adjust baseline Creativity scores, the main effect of the experimental group did not reach statistical significance ($\beta = 0.148$, SEHC3 = $0.134, p = .267$, 95% CI [$-0.113, 0.410$]). This indicates that the unadjusted gain-score finding for Creativity is baseline-sensitive. Detailed sensitivity models and adjusted means are provided in Section 1 of the Supplementary Material.

For the remaining dimensions, such as Self-efficacy, Energy, Risk Propensity, Leadership, and Intent to Innovate, gain differences were not statistically significant after adjustment (all $p_{Holm} > .05$), and their confidence intervals included zero, as detailed in Table 2.

Overall, the lack of significant differential gains across all dimensions after baseline adjustment suggests that the 8-week intervention did not produce robust changes in self-reported innovation competencies, contrasting starkly with the high technical quality observed in the objective performance artifacts. Detailed assumption checks and statistical test justifications are provided in Section 1 of the Supplementary Material.

Table 2. Differential gains in innovation skills (YISMT) comparing experimental and control groups

Dimension	Δ Exp. Mean (SD)	Δ Ctrl. Mean (SD)	Mean Diff.	95% CI [LL, UL]	p_{Holm}	Effect Size (Δ)
Creativity	0.59 (0.97)	0.12 (0.58)	0.47	[0.14, 0.80]	.031*	0.81 (Large)
Self-efficacy	0.33 (1.02)	0.29 (0.70)	0.04	[-0.32, 0.41]	.929	0.06
Energy	0.42 (1.07)	0.26 (0.82)	0.16	[-0.24, 0.57]	.929	0.16
Risk propensity	0.34 (1.00)	0.03 (0.80)	0.31	[-0.08, 0.69]	.741	0.32
Leadership	0.36 (1.18)	0.10 (1.02)	0.26	[-0.22, 0.74]	.916	0.23
Intent to innovate	0.32 (1.20)	0.01 (0.90)	0.31	[-0.13, 0.76]	.830	0.28

Notes: $N = 91$ ($n_{exp} = 61, n_{ctrl} = 30$). CI = Confidence Interval for the difference in means. $p_{Holm} = p$ -value adjusted for multiple comparisons using the Holm-Bonferroni method. Effect sizes are reported as Glass's Δ (using the control group SD as the standardizer) to account for unequal variances. *Significant at $p < .05$.

5 DISCUSSION

This study evaluated the effectiveness of an Ed-CBL intervention for innovation competency development. The results partially confirm the proposed hypotheses, revealing a critical dissociation between technical performance and self-perception: while the intervention generated artifacts of high technical quality validated by the ecosystem (supporting H1), the gains in self-perceived skills were not robust; while an initial unadjusted gain was observed in Creativity, it was nullified after controlling for baseline imbalances (supporting H2). These findings suggest that Ed-CBL architecture functions primarily as a procedural training environment rather than an immediate transformer of stable personality traits.

5.1 The “dosage effect” on creativity: Procedural competence vs. abstract traits

Regarding Hypothesis 2, the initial gain-score analysis identified a significant differential improvement in Creativity (Glass’s $\Delta = 0.81$), with no significant changes in attitudinal dimensions such as Leadership or Self-efficacy (see Table 2). However, although the gain-score analysis showed a significant differential improvement, this result must be interpreted cautiously because Creativity was also the only YISMT dimension showing a significant baseline difference between groups. The baseline-adjusted ANCOVA sensitivity analysis did not confirm the unadjusted Creativity effect, rendering the group effect non-significant. Therefore, the Creativity result should be interpreted as a preliminary, baseline-sensitive finding. While this pattern aligns with recent evidence suggesting that specific innovation skills might be more malleable following structured interventions [41] and corroborates research identifying creativity as a distinct and dynamic factor within the youth innovation profile [42], [43], our adjusted data are not sufficient to establish a robust intervention effect. This underscores the critical need for highly controlled settings to confidently confirm its responsiveness to the “instrumental dose.”

Because the adjusted data do not establish a robust intervention effect on Creativity or other attitudinal traits, the findings validate the psychometric theory that these operate as stable “traits,” requiring incubation periods longer than a short academic cycle [44], [45]. Recent bibliometric and science mapping analyses demonstrate that the intellectual structure of transformational leadership relies on high conceptual maturity and complex, longitudinal evolution [46]. Although technical problem-solving skills (as evidenced by the artifacts) can be acquired rapidly, comprehensive attitudinal frameworks typically stabilize early in life; manifesting detectable shifts therefore requires interventions of significantly greater intensity and duration [47].

These findings establish program duration as a critical predictor of improvement in these complex social outcomes. Rather than indicating a program failure, this stability provides discriminant validity to the instrument; it demonstrates that students did not respond with the generalized inflation bias often observed in educational self-reports and reinforces the study’s core premise: short-term innovation competency is better measured by objective artifact performance than by subjective surveys [45], [48].

5.2 The ecosystem filter: Team attrition as a quality indicator

The attrition analysis revealed that only 55.5% of teams completed the full cycle, further validating Hypothesis 1: the imposition of ecosystem constraints (e.g., SUNASS

regulations) functioned as a selective fidelity filter. The observed attrition (33.3% early dropout) should not be interpreted as mere demotivation but as the activation of “Friction with Reality” [49]. This phenomenon is well-documented in deep-tech ecosystems, where technical complexity generates a natural “selection curve” akin to the “valley of death” in technology startups [50]. Furthermore, engineering education literature suggests that this ambiguity, while pedagogically valuable for creativity, can generate tension and withdrawal if not managed as a deliberate property of instructional design rather than a student failure [51].

The teams completing the full cycle, such as the top-ranked Team 5 (Jury Score: 17.5/20), did not succeed by chance; rather, they effectively pivoted by using technical feedback to refine their artifacts through the quality gates (E1–E4). This process aligns with Design Thinking operational models, demonstrating that final design quality correlates with a team’s capacity to iterate prototypes based on expert feedback during convergence phases [52]. Crucially, this finding challenges the notion that CBL success is measured solely by mass retention; conversely, in the context of “wicked problems,” the completion rate provides information about implementation feasibility and participant progression through the milestone structure [11], [28]. The intervention effectively filtered out technically unfeasible proposals, ensuring that final prototypes met industry standards. External jury validation confirms that the teams completing the full cycle successfully internalized operational constraints, transcending typical academic simulations [4], [8].

5.3 Divergence of measures: Performance vs. self-report

A central methodological contribution of this study lies in the observed discrepancy between the high technical quality of artifacts and the absence of significant gains in perceived self-efficacy. This divergence illustrates the “reference bias” effect: by exposing novice students to rigorous professional standards and expert mentors, their internal frame of reference is recalibrated. Consequently, they become more critical of their own competence, a phenomenon suggesting that self-reports must be weighed cautiously against objective measures [53], [54]. This finding warns against the isolated interpretation of self-efficacy, which may fail to capture actual learning in cognitively demanding environments such as CBL [55]. Moreover, this discrepancy aligns with observations in gamified and technology-enhanced learning environments. Recent literature demonstrates that while short-term instructional novelties can temporarily increase initial engagement or transient enthusiasm, these subjective motivational states do not automatically translate into stable cognitive outcomes or measurable performance traits [56]. Therefore, subjective enthusiasm must not be conflated with objective capability.

Therefore, adhering to the Authentic Assessment principles described in the theoretical framework, we contend that performance evidence (the validated artifact) holds a higher epistemological status than self-report in this context. A team’s ability to deliver a viable water optimization solution (Primary Evidence) demonstrates innovation competency more robustly than subjective survey confidence (Secondary Evidence), aligning with models that prioritize problem-solving in real-world contexts [57]. This methodological triangulation safeguards the study against result inflation based solely on participant satisfaction, a recurring issue in engineering education, and highlights the need for assessment instruments grounded in demonstrable competencies.

5.4 Limitations and future directions

The interpretation of these results must be contextualized within the study's limitations. First, the quasi-experimental design precludes the complete elimination of selection bias, although robust statistical methods (Welch's t-test and Holm's correction) were employed to control for heteroscedasticity [24], [58]. Second, the sample size ($N = 91$) and the single-institution context restrict the immediate generalizability of findings to similar engineering education ecosystems. Although drawing both cohorts from the same applicant pool controlled for baseline motivation, the lack of randomization leaves latent selection biases unmeasured. Future research should strive to employ Randomized Controlled Trials (RCTs) to definitively eliminate these threats to validity.

Third, a specific limitation arising from this non-randomized design concerns the significant baseline imbalance observed in the Creativity dimension between the experimental and control groups. Although the initial gain-score analysis indicated a significant differential improvement, our rigorous sensitivity analysis demonstrated that this main effect did not reach statistical significance after adjusting for baseline scores. Consequently, the findings regarding procedural creativity are explicitly baseline-sensitive and preliminary. This imbalance warrants cautious interpretation and strongly highlights the need to confirm these short-term skill gains in future RCTs or better-matched quasi-experimental designs to definitively eliminate these threats to validity.

Fourth, the measurement of "ecosystem dose" was operationalized as a binary outcome (completed vs. not completed) [27]. Future studies should operationalize stakeholder interaction as a continuous variable (e.g., mentoring minutes) to model more accurate dose-response relationships. Fifth, although the internal conditions of the intervention were rigorously standardized, no prospective follow-up of exogenous variables was carried out in relation to program dropouts. Factors such as students' concurrent academic workloads or external resource constraints could explain this attrition. Future research should incorporate monitoring tools to model attrition determinants (e.g., via logistic regression) and isolate these alternative explanations.

Finally, we recommend extending longitudinal follow-up (3–6 months) to determine if gains in procedural creativity are eventually internalized as stable innovative personality traits. This would allow a transition from evidence of proximal performance to one of sustained behavioral impact [24], [59].

6 CONCLUSION

This study demonstrates that Ed-CBL interventions generate verifiable evidence of engineering competence only when they transcend academic simulation to operate under real ecosystem constraints. The dissociation found between the high technical quality of artifacts (validated by external juries) and the stability in self-perceived skills confirms that "friction with reality" acts as a filter of rigor. Learning, in this context, does not manifest as an inflation of confidence (self-efficacy) but as the operational capacity to navigate technical and regulatory uncertainty. Therefore, it is concluded that the effectiveness of innovative education should not be inferred from student satisfaction but from their progression through external validation gates.

The primary contribution of this study lies in displacing the object of evaluation from “innovation as a latent trait” to “innovation as measurable performance.” By documenting implementation fidelity under TIDieR standards, this study overcomes the limitations of the pedagogical “black box,” establishing that verifiable objective performance (artifact quality) does not necessarily correlate with immediate changes in self-reported competencies. Thus, innovative output in short-term ecosystems is driven by structured, dosed exposure to Design Thinking cycles and expert feedback, even when latent psychological traits remain stable. For engineering institutions, these findings suggest that the scalability of entrepreneurial ecosystems depends less on mass participation and more on the standardization of filtering mechanisms that ensure prototype quality. Future research should focus on longitudinal tracking to determine if these procedural competencies eventually crystallize into stable professional traits.

7 DATA AVAILABILITY AND SUPPLEMENTARY MATERIAL

Supplementary Material for this paper is available online via DOI: [10.17632/bj88dkp4fs.1](https://doi.org/10.17632/bj88dkp4fs.1). Additional data or materials may be requested from the corresponding author.

8 ETHICAL DECLARATION

The program was conducted in accordance with the Declaration of Helsinki and approved under Resolution No. 31376-R-2025. Informed consent was obtained from all subjects involved in the study.

9 REFERENCES

- [1] S. Perna, M. P. Recke, and M. Nichols, “Challenge based learning. A comprehensive survey of literature,” The Challenge Institute, 2023. https://www.challengeinstitute.org/CBL_Literature_Survey.pdf
- [2] J. Larsson, A. Johansson, and O. H. Svensson, “New engineers for a bright future: Deconstructing the promises of challenge based learning,” *Studies in Engineering Education*, vol. 6, no. 1, pp. 129–151, 2025. <https://doi.org/10.21061/see.172>
- [3] A. A. Del Savio, K. G. Velarde, L. C. Montero, and M. A. V. Olivera, “Research training: Unraveling the research methodological design challenge in engineering programs,” *International Journal of Engineering Pedagogy (ijEP)*, vol. 14, no. 7, pp. 30–53, 2024. <https://doi.org/10.3991/ijep.v14i7.48987>
- [4] J. Membrillo-Hernández, M. J. Ramírez-Cadena, M. Martínez-Acosta, E. Cruz-Gómez, E. Muñoz-Díaz, and H. Elizalde, “Challenge based learning: The importance of world-leading companies as training partners,” *International Journal on Interactive Design and Manufacturing*, vol. 13, no. 3, pp. 1103–1113, 2019. <https://doi.org/10.1007/s12008-019-00569-4>
- [5] K. M. Henao-Romero and I. S. E. van der Zande, “Global–local learning ecosystems for societal purpose: A latin American perspective,” *Educ. Sci. (Basel)*, vol. 15, no. 12, p. 1693, 2025. <https://doi.org/10.3390/educsci15121693>
- [6] L. C. Bowling *et al.*, “Addressing water resources and environmental quality programming needs in Arequipa, Peru,” *J. Contemp. Water Res. Educ.*, vol. 173, no. 1, pp. 1–12, 2021. <https://doi.org/10.1111/j.1936-704X.2021.3354.x>

- [7] G. Salmoral *et al.*, “Water-related challenges in nexus governance for sustainable development: Insights from the city of Arequipa, Peru,” *Sci. Total Environ.*, vol. 747, p. 141114, 2020. <https://doi.org/10.1016/j.scitotenv.2020.141114>
- [8] K. Doulougeri, J. D. Vermunt, G. Bombaerts, and M. Bots, “Challenge-based learning implementation in engineering education: A systematic literature review,” *Journal of Engineering Education*, vol. 113, no. 4, pp. 1076–1106, 2024. <https://doi.org/10.1002/jee.20588>
- [9] D. Dunning, K. Johnson, J. Ehrlinger, and J. Kruger, “Why people fail to recognize their own incompetence,” *Curr. Dir. Psychol. Sci.*, vol. 12, no. 3, pp. 83–87, 2003. <https://doi.org/10.1111/1467-8721.01235>
- [10] K. Lavidas, S. Papadakis, D. Manesis, A. S. Grigoriadou, and V. Gialamas, “The effects of social desirability on students’ self-reports in two social contexts: Lectures vs. Lectures and lab classes,” *Information*, vol. 13, no. 10, p. 491, 2022. <https://doi.org/10.3390/info13100491>
- [11] V. Robledo-Rella, L. Neri, R. M. G. García-Castelán, A. Gonzalez-Nucamendi, J. Valverde-Rebaza, and J. Noguez, “A comparative study of a new challenge-based learning model for engineering majors,” *Front. Educ. (Lausanne)*, vol. 10, p. 1545071, 2025. <https://doi.org/10.3389/educ.2025.1545071>
- [12] C. L. Garay-Rondero, A. Castillo-Paz, C. Gijón-Rivera, G. Domínguez-Ramírez, C. Rosales-Torres, and A. Oliart-Ros, “Competency-based assessment tools for engineering higher education: A case study on complex problem-solving,” *Cogent Education*, vol. 11, no. 1, 2024. <https://doi.org/10.1080/2331186X.2024.2392424>
- [13] R. J. Mislevy and G. D. Haertel, “Implications of evidence-centered design for educational testing,” *Educational Measurement: Issues and Practice*, vol. 25, no. 4, pp. 6–20, 2006. <https://doi.org/10.1111/j.1745-3992.2006.00075.x>
- [14] A. Fridrich, G. J. Jenny, and G. F. Bauer, “The context, process, and outcome evaluation model for organisational health interventions,” *Biomed Res. Int.*, vol. 2015, p. 414832, 2015. <https://doi.org/10.1155/2015/414832>
- [15] A. Leles, L. Zaina, and J. Roberto Cardoso, “Challenge-based learning for competency development in engineering education, a prisma-based systematic literature review,” *IEEE Transactions on Education*, vol. 67, no. 5, pp. 746–757, 2024. <https://doi.org/10.1109/TE.2024.3417908>
- [16] K. Helker, M. Bruns, I. M. M. J. Reymen, and J. D. Vermunt, “A framework for capturing student learning in challenge-based learning,” *Active Learning in Higher Education*, vol. 26, no. 1, pp. 213–229, 2025. <https://doi.org/10.1177/14697874241230459>
- [17] J. Höffken and J. Lazendic-Galloway, “Engaging for the future: Challenge-based learning and stakeholder partnerships in sustainability education,” *Sustainable Earth Reviews*, vol. 7, p. 20, 2024. <https://doi.org/10.1186/s42055-024-00087-6>
- [18] T. Van Dang, P. T. G. Bao, V. D. Minh, and N. T. T. Tu, “The application of problem-based learning in soft skills courses: An experiment in classes with multidisciplinary students in Vietnam,” *International Journal of Engineering Pedagogy (IJEP)*, vol. 15, no. 5, pp. 20–41, 2025. <https://doi.org/10.3991/ijep.v15i5.53663>
- [19] G. Zacharis and S. Papadakis, “Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment,” *Educational Process International Journal*, vol. 19, no. 1, 2025. <https://doi.org/10.22521/edupij.2025.19.591>
- [20] L. Fallin and C. Turton, “The transformative potential of design thinking in learning development,” *Journal of Learning Development in Higher Education*, no. 34, 2025. <https://doi.org/10.47408/jldhe.vi34.1314>
- [21] M. Kochanowska and W. R. Gagliardi, “The double diamond model: In pursuit of simplicity and flexibility,” in *Perspectives on Design II*, vol. 16, 2022, pp. 19–32. https://doi.org/10.1007/978-3-030-79879-6_2

- [22] S. Zhang *et al.*, “How to assess AI literacy: Misalignment between self-reported and objective-based measures,” in *LAK26: 16th International Learning Analytics and Knowledge Conference (LAK 2026)*, April 27–May 01, 2026, Bergen, Norway, vol. 1, 2026. <https://doi.org/10.1145/3785022.3785088>
- [23] Q. Meng and Q. Zhang, “The influence of academic self-efficacy on university students’ academic performance: The mediating effect of academic engagement,” *Sustainability (Switzerland)*, vol. 15, no. 7, p. 5767, 2023. <https://doi.org/10.3390/su15075767>
- [24] C. S. Reichardt, “Experimental and quasi-experimental designs for generalized causal inference,” *Social Service Review*, vol. 76, no. 3, pp. 510–514, 2002. <https://doi.org/10.1086/345281>
- [25] A. Rangel-Pacheco and A. Witte, “NeMTSS research brief: Essential components and applications of implementation fidelity,” Nebraska Center for Research on Children, Youth, Families and Schools, 2020. <https://nemtss.unl.edu/wp-content/uploads/2022/06/20-Essential-Components-and-Applications-of-Implementation-Fidelity.pdf>
- [26] L. Simón-Chico, A. González-Peño, E. Hernández-Cuadrado, and E. Franco, “The impact of a challenge-based learning experience in physical education on students’ motivation and engagement,” *Eur. J. Investig. Health Psychol. Educ.*, vol. 13, no. 4, pp. 684–700, 2023. <https://doi.org/10.3390/ejihpe13040052>
- [27] J. F. Jiménez-Parra, N. Belando-Pedreño, J. López-Fernández, A. J. García-Vélez, and A. Valero-Valenzuela, “ACTIVE VALUES: An interdisciplinary educational programme to promote healthy lifestyles and encourage education in values—A rationale and protocol study,” *Applied Sciences*, vol. 12, no. 16, p. 8073, 2022. <https://doi.org/10.3390/app12168073>
- [28] J. Lönngren, “Wicked problems in engineering education: Preparing future engineers to work for sustainability,” *Environ. Educ. Res.*, vol. 25, no. 12, pp. 1808–1809, 2019. <https://doi.org/10.1080/13504622.2019.1639038>
- [29] R. Bender-Salazar, “Design thinking as an effective method for problem-setting and needfinding for entrepreneurial teams addressing wicked problems,” *J. Innov. Entrep.*, vol. 12, p. 24, 2023. <https://doi.org/10.1186/s13731-023-00291-2>
- [30] Sh. M. Hassan, “SCAMPER as a creative idea generation method: Case study on graphic design students,” *Information Sciences Letters*, vol. 12, no. 4, pp. 1417–1426, 2023. <https://doi.org/10.18576/isl/120453>
- [31] N. Nascimento, A. R. Santos, A. Sales, and R. Chanin, “Enablers and inhibitors in Agile Teams – A case study using challenge based learning for mobile application development,” in *Proceedings – 5th International Workshop on Software-Intensive Business: Towards Sustainable Software Business, IWSiB 2022*, 2023, pp. 67–74. <https://doi.org/10.1145/3524614.3528623>
- [32] F. J. Llamas Fernández and J. C. Fernández Rodríguez, “La metodología Lean startup: desarrollo y aplicación para el emprendimiento,” *Revista Escuela de Administración de Negocios*, no. 84, 2018. <https://doi.org/10.21158/01208160.n84.2018.1918>
- [33] E. A. López-Guajardo, R. A. Ramirez-Mendoza, A. Vargas-Martinez, W. Jianhong, A. Roman-Flores, and G. Zavala, “Argumentative-driven assessments in engineering: A challenge-based learning approach to the evaluation of competencies,” *International Journal on Interactive Design and Manufacturing*, vol. 17, no. 1, pp. 79–91, 2023. <https://doi.org/10.1007/s12008-022-01188-2>
- [34] S.-H. Jin, K.-I. Song, D. Shin, and S. Shin, “A performance-based evaluation rubric for assessing and enhancing engineering design skills in introductory engineering design courses,” *International Journal of Engineering Education*, vol. 31, pp. 1007–1020, 2015.
- [35] P. E. Shrout and J. L. Fleiss, “Intraclass correlations: Uses in assessing rater reliability,” *Psychol. Bull.*, vol. 86, no. 2, pp. 420–428, 1979. <https://doi.org/10.1037//0033-2909.86.2.420>

- [36] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, 2016. <https://doi.org/10.1016/j.jcm.2016.02.012>
- [37] E. Chell and R. Athayde, "The identification and measurement of innovative characteristics of young people: Development of the youth innovation skills measurement tool," National Endowment for Science, Technology and the Arts, 2009. [Online]. Available: <https://researchke.kingston.ac.uk/ws/portalfiles/portal/48219303/Chell-E-5985.pdf>
- [38] B. G. Tabachnick and L. S. Fidell, "Using multivariate statistics," *Journal of Applied Psychology*, vol. 87, no. 4, pp. 611–628, 2017. [Online]. Available: <https://lccn.loc.gov/2017040173>
- [39] P. D. Ellis, "The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results," in *The Essential Guide to Effect Sizes*, 2010. <https://doi.org/10.1017/CBO9780511761676>
- [40] S. E. Maxwell, H. D. Delaney, and K. Kelley, "Designing experiments and analyzing data: A model comparison perspective, third edition," in *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, Third Edition, 2017, pp. 1–1056. <https://doi.org/10.4324/9781315642956>
- [41] S. Gupta, "Can innovation be taught? Evidence from a school-based RCT in India," Stanford | King Center of Global Development, 2025. <https://kingcenter.stanford.edu/sites/g/files/sbiybj16611/files/media/file/wp2076.pdf>
- [42] G. Özdemir Aydın, S. Küçükaydınoğlu, A. Palloş, E. Y. Coşkun, and N. Turan, "The relationship between first-year nursing students' innovation skills and attitudes toward artificial intelligence: A multicentre study," *Nurse Educ. Pract.*, vol. 86, p. 104429, 2025. <https://doi.org/10.1016/j.nepr.2025.104429>
- [43] H. Z. Altınışık, T. Adıgüzel, and Y. G. Gençer, "Adaptation of youth innovational skills measurement tool for Turkish usage," *Kastamonu Education Journal*, vol. 31, no. 1, pp. 155–164, 2023. <https://doi.org/10.24106/kefdergi.1246467>
- [44] D. L. Head, "A novel game-based assessment of risk-taking propensity: Development and psychometric evidence," UNLV Theses, Dissertations, Professional Papers, and Capstones, 2025. <https://doi.org/10.34917/39206727>
- [45] B. Lira *et al.*, "Large studies reveal how reference bias limits policy applications of self-report measures," *Sci. Rep.*, vol. 12, no. 1, p. 19189, 2022. <https://doi.org/10.1038/s41598-022-23373-9>
- [46] T. Karakose, T. Tülübaş, S. Papadakis, and R. Yirci, "Evaluating the intellectual structure of the knowledge base on transformational school leadership: A bibliometric and science mapping analysis," *Education Sciences*, vol. 13, no. 7, p. 708, 2023. <https://doi.org/10.3390/educsci13070708>
- [47] J. P. Doh, "Can leadership be taught? Perspectives from management educators," *Academy of Management Learning & Education*, vol. 2, no. 1, pp. 54–67, 2017. <https://doi.org/10.5465/amle.2003.9324025>
- [48] A. L. Duckworth and D. S. Yeager, "Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes," *Educ. Res.*, vol. 44, no. 4, p. 237, 2015. <https://doi.org/10.3102/0013189X15584327>
- [49] P. Pantzos, A. Pears, and E. Ampadu, "Students' motivation in challenge-based learning in higher engineering education: A scoping review," *International Journal of Engineering Pedagogy (IJEP)*, vol. 15, no. 6, pp. 4–34, 2025. <https://doi.org/10.3991/ijep.v15i6.54571>
- [50] A. G. L. Romme, "Against all odds: How eindhoven emerged as a deeptech ecosystem," *Systems*, vol. 10, no. 4, p. 119, 2022. <https://doi.org/10.3390/systems10040119>
- [51] C. Fang, G. R. Dantes, K. R. Dantes, and D. S. Wahyuni, "Design-based learning in undergraduate engineering education: A systematic conceptual review of implementation, impact, and challenges," *PPSDP International Journal of Education*, vol. 4, no. 2, pp. 1–16, 2025. <https://doi.org/10.59175/pijed.v4i2.774>

- [52] E. Y. L. Au and R. S. Goonetilleke, "Leveraging design thinking to enhance engineering teaching: An operational model," *IEEE Transactions on Education*, vol. 68, no. 1, pp. 95–102, 2025. <https://doi.org/10.1109/TE.2024.3467387>
- [53] J. Power and D. Tanner, "Peer assessment, self-assessment, and resultant feedback: An examination of feasibility and reliability," *European Journal of Engineering Education*, vol. 48, no. 4, pp. 615–628, 2023. <https://doi.org/10.1080/03043797.2023.2185769>
- [54] P. Dawson, "Assessment rubrics: Towards clearer and more replicable design, research and practice," *Assess. Eval. High. Educ.*, vol. 42, no. 3, pp. 347–360, 2017. <https://doi.org/10.1080/02602938.2015.1111294>
- [55] S. I. Marino, "Enhancing graduation outcomes through challenge-based learning: A strategic proposal," *J. Educ. Elearn. Res.*, vol. 12, no. 3, pp. 499–506, 2025. <https://doi.org/10.20448/jeelr.v12i3.7397>
- [56] A. I. Zourmpakis, M. Kalogiannakis, and S. Papadakis, "Adaptive gamification in science education: An analysis of the impact of implementation and adapted game elements on students' motivation," *Computers*, vol. 12, no. 7, p. 143, 2023. <https://doi.org/10.3390/computers12070143>
- [57] J. A. Palma-Mendoza, D. E. Salinas-Navarro, I. A. Arana-Solares, and F. Franco-Herrera, "Challenge-based learning for active learning in industrial engineering education," in *The Emerald Handbook of Active Learning For Authentic Assessment*, 2025, pp. 313–335. <https://doi.org/10.1108/978-1-83797-857-120251016>
- [58] Y.-C. Ching and Y.-C. Ho, "AI virtual human-augmented game-based teaching to enhance emotional intelligence in nursing students: Protocol for a single-group pretest-posttest action research study," *JMIR Res. Protoc.*, vol. 14, p. e80290, 2025. <https://doi.org/10.2196/80290>
- [59] S. Sulaiman and K. Z. Mamaraimovna, "Enhancing professional competence through ESP," *American Journal of Philological Sciences*, vol. 5, no. 6, pp. 182–185, 2025. <https://doi.org/10.37547/ajps/Volume05Issue06-49>

10 APPENDIX

10.1 Appendix A: Intervention fidelity

Table A1. Template for intervention description and replication (TIDieR) checklist

Item (Descriptor)	Details (Derived from Innovation Challenge 2025 Protocol)
Brief Name Name	Ed-CBL Innovation Challenge 2025
Why Rationale	To develop innovation competencies through a "wicked problem" anchor: Optimization of drinking water in scarcity zones (SDG 6), subject to real-world industry constraints imposed by SEDAPAL/SUNASS.
What Materials	LMS Platform: EQUIPU. Design Thinking Templates: Problem Tree, User Persona, Experiment Canvas, MoSCoW Matrix. Guidelines: Standardized Mentoring Protocol.
What Procedures	Fixed Sequence: Kick-off, Welcome Session, 4 Workshop Cycles (E1–E4), and Local Pitch. Each cycle included a synchronous workshop followed by a mandatory "gatekeeping" mentoring checkpoint.

(Continued)

Table A1. Template for intervention description and replication (TIDieR) checklist (*Continued*)

Item (Descriptor)	Details (Derived from Innovation Challenge 2025 Protocol)
Who Provider	Support: 18 Local Faculty Mentors Validation: External experts from SEDAPAL and SUNASS (Technical Advisory).
How Mode	Hybrid: Synchronous milestones via Zoom; asynchronous teamwork; technical coordination via WhatsApp channels (Mentors/Participants groups).
Where Setting	Institutional node in Arequipa, Peru, extending to peri-urban and high-Andean rural zones for field validation.
When Dosing	Duration: March 25–June 20, 2025 (8 weeks). Intensity: 1 Kick-off (2h), 1 Welcome (3h), 4 Workshops (4h each), and 4 Mentoring Sessions (1h per team/milestone).
Tailoring Adaptation	Multidisciplinary requirement: Teams of 3–5 members representing ≥ 3 distinct academic schools.
Modifications Changes	No formal protocol amendments reported. Ad-hoc deadline extensions were logged in to the official communication channel.
Planned fidelity Strategies	Fidelity ensured via: Mandatory deliverables, mentoring as quality gates, and standardized rubrics (see Appendix B).
Actual fidelity Adherence	Completion Rate: 15/27 teams (55.5%) completed the full cycle. 12 teams dropped out or failed gatekeeping checks.

Note: Adapted from the TIDieR checklist to ensure replicability of the Ed-CBL intervention.

10.2 Appendix B: Assessment rubrics

Table B1. Gatekeeping rubric for intermediate deliverables (process) used for milestones E1–E4

Criterion	1 Insufficient	3 Acceptable	5 Excellent
1. Customer Journey Map (CJM)	Vague persona/ journey; missing stages or emotions; confusing diagram.	Defined persona; CJM with basic stages and some pain points; correct visualization.	Validated Persona: Data-backed profile; detailed CJM (moments of truth, KPIs); actionable insights derived from field validation.
2. Ideation (SCAMPER)	Few generic ideas; no connection to CJM insights.	Relevant ideas: moderate creativity linked to some insights.	Disruptive and Viable: Numerous ideas; rigorous application of SCAMPER verbs to key insights; clear value justification.
3. Feasibility Filter (MoSCoW)	Confused list; requirements misclassified without criteria.	Logically organized requirements; basic alignment with user needs.	Rigorous Prioritization: Argumentation based on user data and technical feasibility; clear traceability to CJM and regulatory constraints.
4. MVP Canvas	Incomplete canvas; diffuse hypotheses; undefined value proposition.	Coherent canvas; clear backlog and alternative hypotheses.	Robust Canvas: Validated segment, concrete value proposition, initial success metrics defined; alignment with SDG 6 targets.

Notes: Adapted from the Innovation Challenge 2025 assessment protocol. Scores of 2 and 4 were assigned for performance levels falling between the criteria described. This rubric was applied by mentors at each gatekeeping milestone to authorize progression to the next phase.

Table B2. Final pitch evaluation rubric (summative) used by external jury

Criterion (Weight)	Description of High Performance (5 Points)
Innovation and Solution Focus (20%)	Disruptive Proposal: Generates high community value by significantly improving upon existing processes, products, or services.
Problem Identification (context) (20%)	Critical Analysis: Clearly delimits a specific problem based on social/geographic context; variables explicitly define the needs of the target audience.
Economic Viability and Sustainability (20%)	Resource Strategy: Proposals define necessary execution resources, value creation scenarios, and investment strategies to ensure long-term sustainability.
Social Impact (15%)	Systemic Change: Evidence that the solution addresses root needs of community actors, supported by rigorous field research.
Scalability (15%)	Growth Potential: Demonstrates capacity for exponential replication in other communities (e.g., nationwide) based on identified commonalities.
Pitch Presentation (10%)	Storytelling: Coherent synthesis of problem/solution; demonstrates team agency (“empowerment”) and clarity of operation from start to finish.

Notes: This rubric was applied by the external jury panel consisting of experts from the ANA, SEDAPAR, and industry specialists. Raw scores for each criterion were assessed on a 0–5 point scale and subsequently weighted according to the specified percentages to calculate a final aggregated score on a 0–20 scale.

11 AUTHORS

Dely Lazo-Barreda is affiliated with the School of Systems Engineering, Universidad Católica de Santa María, Arequipa, Peru. She holds a Master of Business Administration and a Master of Information Technology Management. Her academic and professional background is related to systems engineering, business administration, and information technology management (E-mail: dlazo@ucsm.edu.pe).

Elizabeth-Susan Mamani-Machaca is affiliated with Universidad San Ignacio de Loyola, Peru. She is a Commercial Engineer with a specialization in Finance, a Registered Professional Engineer, and a RENACYT-certified researcher. She has experience in university quality management, social responsibility, institutional planning, and data analytics. Her research interests include research management, organizational sustainability, integrated management systems, and SDG-oriented social innovation (E-mail: elizabeth.mamani@epg.usil.pe).

Karina Rosas-Paredes is affiliated with Universidad Católica de Santa María, Arequipa, Peru. She holds a PhD in Systems Engineering with a focus on Information Security and serves as Director of Innovation and Development of the Vice-Rectorate for Research at UCSM. Her research interests include cybersecurity, education, STEM, innovation, and entrepreneurship (E-mail: kparedes@ucsm.edu.pe).

Midwar Ancco is affiliated with Universidad Católica de Santa María, Arequipa, Peru. He holds a Master of Science and has experience in R&D&I projects, with training as a Biotechnology Engineer from Universidad Católica de Santa María. He specializes in industrial and environmental biotechnology and currently serves as Teaching Practice Coordinator at UCSM (E-mail: manccom@ucsm.edu.pe).

Carla Cuya-Zevallos is affiliated with the Faculty of Nursing, Universidad Católica de Santa María, Arequipa, Peru. She holds a PhD in Public Health and a Master of Public Health. She is an ICU nurse specialist and faculty member at the Faculty of Nursing, Universidad Católica de Santa María, Arequipa, Peru (E-mail: ccuya@ucsm.edu.pe).