Assessing the perceived effectiveness of a design thinking-based sustainability simulator: An exploratory higher-education study

Cinthya Flores-Rivera[1]; César Lugo-Medrano[2]; María Yolanda Burgos-López[3]
[1,3] Tecnologico de Monterrey, México, *cinthya.flores@tec.mx, yolanda.burgos@tec.mx*
[2]Tecnologico de Monterry, México, *cesar.lugo@tec.mx*

Abstract—This study explores how a Design Thinking-oriented educational intervention can be evaluated when simulation is used to teach sustainability-related concepts. The analysis uses 73 post-intervention activity responses contributed by 66 higher-education students, seven of whom completed two distinct simulator activities. The study develops PEDTS, a 0–100 exploratory composite index integrating learning experience, meaningful learning, enjoyment, and engagement. Internal convergence, participant-clustered inference, one-observation-per-participant sensitivity analysis, association with overall user experience, and permutation-based profile comparisons were examined. PEDTS remained substantively stable in the 66-participant sensitivity sample (M=87.39; α=0.935) and retained a large association with UX (ρ=0.832). Participant-clustered inference confirmed the activity-level association (β=0.739, p<.001). The findings support PEDTS as a concise exploratory evaluation layer for Design Thinking-oriented sustainability simulations while maintaining a clear distinction between perceived effectiveness and objective sustainability or entrepreneurial outcomes.

Keywords—Design thinking; sustainability education; simulation-based learning; educational innovation; user experience

I. INTRODUCTION

Sustainability education requires learning environments in which students can interpret interconnected problems, consider competing priorities, and explore the consequences of decisions. Simulation-based learning is appropriate for this purpose because it moves students from recalling concepts toward acting within a decision environment. Previous research has shown that simulation can improve sustainability understanding and can provide some transfer to structurally similar problems [1]. Recent work also describes sustainability simulation as a bridge between theoretical content and practical decision-making in environmental, economic, and policy contexts [2].

Design Thinking (DT) adds a human-centered and iterative logic. Students examine stakeholder needs, frame a challenge, generate alternatives, make ideas tangible, obtain feedback, and refine decisions. In sustainability education, this approach can help transform broad goals into situated challenges involving uncertainty, trade-offs, and multiple forms of value. Higher-education research has identified DT practices that can support sustainability values, engagement with complexity, and the capacity to envision sustainable futures [3].

Although simulation-based learning has been widely used in sustainability education, most studies evaluate educational interventions through isolated constructs such as engagement, satisfaction, usability, or perceived learning. Consequently, instructors often lack an integrated indicator capable of summarizing the overall educational value of a Design Thinking-oriented simulation while preserving its cognitive and experiential dimensions. This gap limits comparisons across interventions and makes it difficult to monitor educational effectiveness through a concise yet theoretically grounded measure.

Accordingly, this study develops and examines the Perceived Effectiveness of the Design-Thinking-Based Sustainability Simulator (PEDTS), an exploratory composite educational indicator intended to summarize perceived effectiveness rather than a fully validated psychometric scale.

Research question: To what extent does PEDTS provide coherent and meaningful evidence for evaluating a Design Thinking-oriented sustainability simulation? Can PEDTS discriminate among different types of student experiences during the simulation?

- *H1 states that PEDTS is positively associated with students' overall user experience. A complementary exploratory expectation is that PEDTS differs across qualitative experience profiles.*

The sustainability problem addressed in this study revolves around fostering students' understanding and application of sustainable practices in real-world contexts. The simulation utilized is called "SomosRetec.com", an educational platform designed to immerse students in a structured challenge aimed at

promoting sustainability knowledge and decision-making skills.

This study was conducted with students enrolled in the course "Design and Analysis of Experiments for Engineering Innovation" from two different groups: one based at the Monterrey campus and the other at the Sinaloa campus. Through the "SomosRetec.com" simulator, students participated in a week-by-week challenge that required them to complete various activities aligned with sustainability practices. The objective was to deepen their comprehension of sustainability principles while fostering actionable learning. As they progressed, students earned badges to recognize their achievements and milestones in meeting the challenge's goals. These badges served as motivational tools, promoting engagement and reinforcing their commitment to sustainability education.

By integrating the "SomosRetec.com" simulator with a Design Thinking approach, students experienced a dynamic learning environment that emphasized iterative problem-solving, stakeholder consideration, and the exploration of trade-offs. This study examines how such an approach, combined with the PEDTS framework, can provide meaningful evidence for evaluating the perceived effectiveness of sustainability-focused educational simulations.

II. THEORETICAL FRAMEWORK

A. Design Thinking as an Educational Approach

A common DT representation includes empathizing, defining, ideating, prototyping, and testing. The phases are iterative rather than strictly linear. Empathy directs attention to stakeholders; definition converts observations into a workable challenge; ideation expands the solution space; prototyping converts possibilities into discussable alternatives; and testing provides evidence for revision. The educational value lies in repeated movement between understanding, experimentation, and improvement.

For sustainability, DT supports challenge-centered learning and the integration of knowledge. The DT4S initiative explicitly connects DT with problem solving, collaboration, and active engagement with both challenge parameters and affected individuals [4]. Sustainability-oriented DT can therefore provide a structure for exploring complex needs and possible futures rather than presenting sustainability as a list of isolated principles [3].

DT is also relevant to entrepreneurship education because opportunity recognition, stakeholder understanding, experimentation, and responsible value creation require similar capabilities. A social-entrepreneurship program integrating DT reported pre-post improvements in entrepreneurial competencies [5]. In the present study, entrepreneurship is an educational implication, not a measured outcome, because the original questionnaire did not include an entrepreneurial-competency scale.

B. Innovation, Sustainability, and Simulation

Simulation can make system relationships visible, allow repeated decisions, and expose learners to the consequences of alternative choices. Research combining systems thinking and simulation reported gains in understanding a sustainability problem [1]. Design-based research has also emphasized simulation as a practical environment for environmental management, economic planning, and policy decisions [2][12].

A positive simulation experience, however, is not identical to sustainability competence. PEDTS addresses the perceived effectiveness of the learning experience. Direct sustainability measures remain necessary to determine whether students improved systems thinking or selected more responsible alternatives. This layered evaluation is useful: PEDTS indicates whether the intervention is educationally engaging and meaningful, while decision metrics establish whether the intervention changed sustainable performance.

C. PEDTS and Hypothesis

The four dimensions were selected a priori because they represent two complementary domains discussed in educational-experience research. Learning Experience and Meaningful Learning capture perceived cognitive-educational value, whereas Enjoyment and Engagement represent experiential quality. The indicators were adapted conceptually from meaningful-learning theory [8], UX research emphasizing pragmatic and hedonic quality [7], human-centered design guidance [9], and emotional-design principles [6]. Because each indicator was administered as a single 0–100 rating, PEDTS is treated as an exploratory composite index and not as a new multi-item psychometric instrument.

PEDTS is the unweighted mean of four equally scaled indicators: Equal weighting was specified a priori because all indicators share the same metric, each domain is theoretically necessary, and no defensible basis existed for privileging one component. This transparent specification avoids sample-specific weights and facilitates replication.

$$PEDTS_i = (\text{Learning Experience}_i + \text{Meaningful Learning}_i + \text{Enjoyment}_i + \text{Engagement}_i) / 4$$

Overall UX is deliberately excluded from the formula and retained as an external criterion. PEDTS therefore summarizes students' integrated appraisal of cognitive and experiential value without embedding the validation outcome inside the index.

H1: PEDTS is positively associated with students' overall user experience.

III. METHODOLOGY

A. Design and Participants

The study used a quantitative, cross-sectional, post-intervention design. The database contained 73 complete activity-level responses contributed by 66 unique students. Participants were 17–21 years old (M=18.70); 36 responses were recorded as

female and 37 as male. Thirty-two responses corresponded to first-semester students and 41 to third-semester students. All records indicated informed consent. The completed simulator activity was the primary observational unit; because seven students contributed responses to two different activities, regression inference used participant-clustered standard errors and all analyses were repeated using one chronologically first response per student (n=66).

The seven repeated identifiers represented valid participation in two different simulator activities rather than duplicate data entry. Dependence was addressed by clustering regression standard errors at the participant level and by a deterministic sensitivity analysis retaining the first chronologically recorded activity for each student. The two approaches preserve the activity-level information while testing whether repeated participation changes the substantive conclusions.

B. Simulator, Sustainability Context, and Implementation

The simulator utilized in this study is "SomosRetec.com", an educational platform developed to promote sustainability knowledge and decision-making skills. The simulator interface is user-friendly and interactive, guiding students through a structured challenge related to sustainability practices. The sustainability topic addressed was the implementation of clean technologies, with a focus on understanding and balancing environmental, social, and economic trade-offs.

The scenario engaged students in a series of weekly activities designed to simulate real-world sustainability challenges over a four-week period. Working in teams, students dedicated approximately five hours per week to these tasks, making critical decisions regarding resource allocation, technology adoption, and long-term impact evaluation. Real-time feedback allowed them to observe the immediate consequences of their choices and fostered reflective learning, while a badge-based system recognized key achievements and sustained student motivation.

C. Measures

Students assessed Learning Experience, Meaningful Learning, Enjoyment, Engagement, and Overall User Experience (UX) using 0–100 scales (0="Not at all," 100="Extremely"). PEDTS was calculated as the average of the first four variables, while UX served as a separate validation criterion.

Questionnaire Wording and Scale Anchors:

- Learning Experience: "To what extent did this activity contribute to your learning?"
- Meaningful Learning: "How meaningful or relevant was the content?"
- Enjoyment: "How much did you enjoy the activity?"
- Engagement: "To what extent did you feel engaged?"
- Overall UX: "How would you rate your overall experience?"

Students also selected one final experience descriptor from predefined response options. Four analytical profiles were generated through the a priori deterministic mapping used in the original study: High Emotional Value (Happy, Pleasant, Inspiring, Relevant, Deep), High Cognitive Demand (Challenging, Stressful, Exhausting), Emotional Rejection (Boring, Irritating, Annoying, Demotivating), and Low Perceived Value (Irrelevant, Useless, Superficial). Because classification applied a fixed dictionary rather than interpretive coding of open-ended narratives, intercoder reliability was not applicable. The mapping is reported to ensure reproducibility.

Administration Mode: The questionnaire was administered online via SomosRetec.com immediately after the final session.

Source and measurement status: The single-item ratings were adapted conceptually from established educational-learning and UX traditions [6]–[9] and tailored to the simulator context. The five questions do not reproduce any complete published scale. They provide concise activity-level indicators, and PEDTS is evaluated as an exploratory composite index.

D. Analysis

Validation combined complementary evidence. Cronbach's alpha, corrected item-total correlations, KMO, Bartlett's test, and PCA examined internal convergence but were not interpreted as full psychometric validation. Association with UX was treated as convergent or criterion-related evidence because the measures were collected from the same respondent at the same time and are conceptually adjacent. Spearman correlation described the bivariate association. OLS inference used participant-clustered standard errors. A sensitivity analysis retained one chronologically first response per student (n=66) and repeated descriptive, convergence, association, and profile analyses. Profile differences were evaluated with Kruskal–Wallis H and a 10,000-replication Monte Carlo permutation test because two profiles were very small.

IV. RESULTS

A. Overall Perceived Effectiveness

PEDTS averaged 87.62 points (SD=15.26), with a median of 92.50 and range 24.50–100.00 across 73 activity responses. In the 66-participant sensitivity sample, the corresponding estimates were M=87.39, SD=15.82, median=92.50, and range 24.50–100.00. The close agreement indicates that repeated participation did not drive the overall level or dispersion of PEDTS.

The distribution was non-normal (Shapiro–Wilk $W = .780$, $p < .001$) and concentrated near the upper boundary. This pattern is consistent with a generally successful intervention, but it also supports the decision to use rank-based and robust methods. For practical reporting, 60 responses (82.2%) scored at least 80. These ranges are descriptive communication aids rather than validated diagnostic cutoffs.

The shape of the distribution also carries design information. A high central tendency suggests that the intervention generated value across cognitive and experiential dimensions simultaneously, rather than succeeding only as an entertaining activity or only as a demanding academic task. At the same time, the lower-scoring tail indicates that the experience was not equally effective for every participant. From a DT perspective, these cases are diagnostically important because they identify users for whom the educational solution may require reframing, additional scaffolding, or a different feedback sequence. PEDTS therefore functions not only as a summary score but also as an instrument for locating heterogeneous user responses that can feed the next redesign cycle.

TABLE I. DESCRIPTIVE RESULTS FOR PEDTS

Statistic	Value
N	73
Mean (SD)	87.62 (15.26)
Median	92.50
Observed range	24.50–100
Distribution	W=.780, p<.001

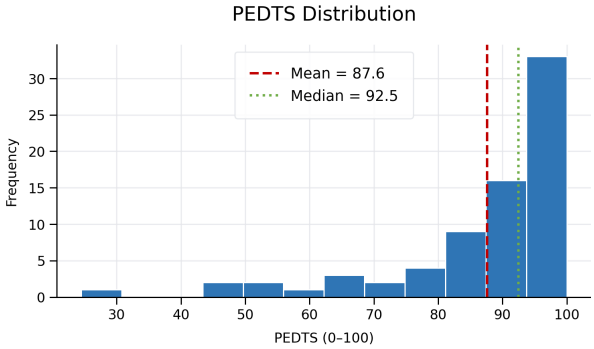


Fig. 1. Distribution of PEDTS scores.

B. Coherence of the PEDTS Components

The four components showed substantial internal convergence in the activity-level data (α=0.932) and in the participant-level sensitivity sample (α=0.935). These coefficients summarize convergence among four conceptually related single-item indicators; they are not interpreted as full reliability or psychometric validation of a multi-item instrument.

Exploratory internal-structure results were also stable. For the 73 activity responses, KMO=0.747, Bartlett $\chi^2(6)=282.83$, p<.001, and the first principal component explained 83.93% of variance with loadings 0.902–0.960. For 66 unique participants, KMO=0.735, Bartlett $\chi^2(6)=269.83$, p<.001, and the first component explained 84.71% with loadings 0.906–0.967. The replicated one-component pattern supports PEDTS as a parsimonious composite summary in this dataset, not as a fully validated psychometric scale.

The convergence is theoretically meaningful. Learning experience and meaningful learning represent whether the activity was perceived as pedagogically valuable, whereas enjoyment and engagement represent whether students were willing and able to remain involved in the experience. Their joint behavior suggests that the perceived effectiveness of the simulator was not fragmented into an academic side and an experiential side. Instead, respondents tended to evaluate both sides consistently. This integrated pattern is compatible with the user-centered premise of DT: an educational solution is stronger when it supports the intended learning purpose while remaining usable, engaging, and worth continuing.

The reliability evidence should nevertheless be interpreted as support for an exploratory composite indicator rather than as proof that four single ratings constitute a definitive psychometric scale. The important result for the present study is that equal weighting did not combine unrelated measures. Each component contributed strongly, and no component appeared dispensable. This finding supports the parsimonious PEDTS calculation used in the subsequent analyses.

TABLE II. RELIABILITY AND INTERNAL-STRUCTURE EVIDENCE

Evidence	Result	Meaning
Cronbach's α	.932	High coherence
Item-total r	.807–.882	All components contribute
KMO	.747	Adequate common variance
PC1 variance	83.93%	Dominant general appraisal
PC1 loadings	.902–.960	Strong convergence

C. Relationship with Overall UX

PEDTS showed a large positive association with overall UX in the activity-level data (Spearman ρ=0.836, p<.001) and in the 66-participant sensitivity sample (ρ=0.832, p<.001). This is interpreted as convergent or criterion-related association rather than strong external validation because PEDTS and UX were provided by the same respondent at the same time and are conceptually adjacent.

The activity-level OLS coefficient was β=0.739 (R^2=0.617). With standard errors clustered by participant, the coefficient remained reliable, SE=0.168, 95% CI [0.410, 1.068], p=1.07e-05. The 66-participant sensitivity analysis yielded β=0.732 and R^2=0.619. Direction and magnitude were therefore stable after accounting for within-student dependence, supporting H1 as an association.

The magnitude is educationally relevant but should be interpreted with conceptual proximity and shared method in mind. PEDTS tracks a related global appraisal; it does not establish discriminant validity from UX or a causal effect. Overall UX may additionally capture interface clarity, aesthetics, navigation, accessibility, and technical stability, which are not represented in PEDTS.

For instructors and designers, the coefficient translates the statistical result into a design proposition. Improvements that increase meaningful learning, learning experience, enjoyment, or engagement together are likely to accompany a more favorable overall evaluation. The result does not indicate which

component should be changed first in every context. Component-level inspection is still required to identify the weakest element and determine whether the next iteration should focus on pedagogical structure, challenge calibration, feedback, or experiential design.

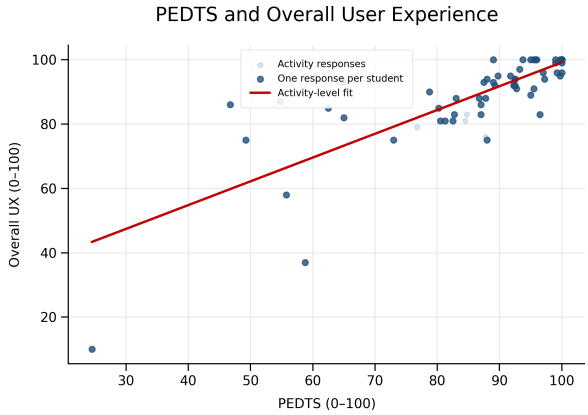


Fig. 2. Association between PEDTS and overall user experience.

D. Experience Profiles and Productive Challenge

PEDTS varied across the four prespecified profiles. In the 73 activity responses, Kruskal–Wallis H=12.38; the 10,000-replication permutation p value was 0.0020. In the 66-participant sensitivity sample, H=11.32 and permutation p=0.0039. The high-emotional-value profile recorded the highest mean (91.01), followed by high cognitive demand (83.23), low perceived value (66.19), and emotional rejection (56.00) in the activity-level data. The permutation and sensitivity results support an exploratory ordered pattern, while the very small low-value and rejection groups preclude definitive pairwise claims.

The cognitive-demand result is especially relevant to DT. A demanding experience was not automatically an ineffective one. Iteration, ambiguity, and the comparison of competing solutions can require effort while still producing meaningful learning and engagement. However, the original grouping combined "challenging" with "stressful" and "exhausting." Future evaluations should distinguish productive challenge from overload so that instructors can preserve intellectual difficulty without allowing unnecessary friction to dominate the experience.

The approximately eight-point difference between the high-emotional-value and high-cognitive-demand means suggests that positive affect and perceived effectiveness are related but not identical. Students may recognize substantial educational value even when the activity requires considerable effort. By contrast, the lower means for low perceived value and emotional rejection indicate that PEDTS is sensitive to experiences in which cognitive and experiential benefits are not jointly realized. This ordered gradient provides an interpretive bridge between the numerical index and the language students used to characterize the activity.

These profile results are treated as pattern evidence rather than a ranking of student types. The two smallest activity-level profiles contained four and two observations. Accordingly, no post-hoc pairwise significance claims are made. A larger replication should prespecify contrasts between productive challenge, overload, low value, and rejection and collect short narrative explanations of the relevant simulator features.

TABLE III. PEDTS BY FINAL-EXPERIENCE PROFILE

Experience profile	n	Mean PEDTS
High emotional value	57	91.01
High cognitive demand	10	83.23
Low perceived value	4	66.19
Emotional rejection	2	56.00

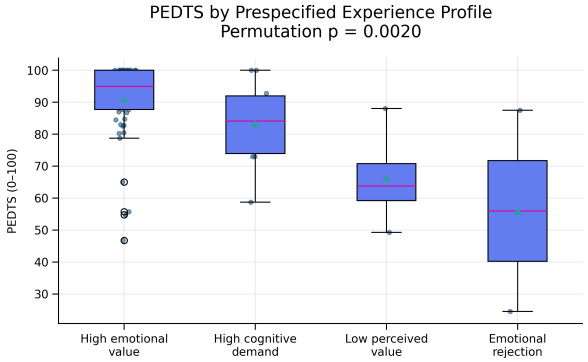


Fig. 3. PEDTS by qualitative final-experience profile.

V. DISCUSSION

A. Contribution to Design Thinking and Sustainability Education

The results suggest that students evaluated the intervention holistically. Learning, meaningfulness, enjoyment, and engagement converged into a strong overall appraisal rather than operating as disconnected reactions. This complements DT-for-sustainability research emphasizing challenge-centered, collaborative, and future-oriented learning [3], [4].

Unlike previous studies that analyze engagement, learning or enjoyment separately, PEDTS offers an integrated educational dashboard capable of summarizing multiple dimensions of perceived effectiveness into a single interpretable score while preserving access to each underlying component.

The result is also compatible with sustainability-simulation research showing that interactive tools can support understanding and practical engagement [1], [2]. This study contributes a different form of evidence: it evaluates perceived effectiveness rather than objective knowledge or transfer. PEDTS should therefore accompany direct outcome measures. Used in this way, the index can help explain why a technically sound simulator succeeds or struggles as a learning experience.

The contribution can be positioned as a meso-level evaluation framework between process fidelity and learning outcomes. Process fidelity asks whether the DT phases and simulator activities were implemented as intended. Outcome assessment asks whether sustainability knowledge, entrepreneurial

capability, or decision quality improved. PEDTS addresses the intermediate question of whether students experienced the intervention as simultaneously educational, meaningful, enjoyable, and engaging. This middle layer can help explain why the same instructional design produces different outcomes across cohorts or contexts.

B. Implications for Entrepreneurial Skills and Sustainable Decisions

DT-oriented simulation can create conditions relevant to entrepreneurial learning: understanding stakeholders, framing opportunities, comparing alternatives, experimenting, and refining value propositions [5] [11]. PEDTS indicates whether students perceived that environment as educationally valuable and engaging. Instructors can combine PEDTS with an entrepreneurial-output rubric covering opportunity quality, stakeholder value, feasibility, experimentation, and responsible value creation [5][11]. This would connect experience quality with the actual quality of student proposals [6][7].

For sustainable decision-making, future implementations should capture the alternatives selected, environmental and social consequences, economic trade-offs, response to feedback, and revisions between simulation rounds [6][10]. Those data would permit a direct test of whether students with higher PEDTS also make more balanced or responsible decisions. The present study therefore establishes an evaluation foundation while specifying the behavioral evidence needed for the next stage.

C. Educational Innovation and Practical Use

PEDTS is simple to calculate and remains on the familiar 0–100 scale. Because it combines cognitive and experiential value, it can serve as a dashboard indicator for comparing cohorts, simulator versions, or instructional designs. Component-level scores should still be retained. Two cohorts may have the same PEDTS mean but differ in whether improvement is needed in meaningful learning, enjoyment, or engagement. The overall index supports monitoring, while the components guide redesign.

D. Future Research Agenda

The results support an ambitious research agenda in which PEDTS becomes one layer of a multilevel model of DT-supported sustainability education. The next stage should test not only whether learners value the experience, but also how that value is generated and when it translates into entrepreneurial and sustainability outcomes.

First, future studies should use a longitudinal or quasi-experimental design with pre-intervention, post-intervention, and delayed measurement. A comparison condition without the DT sequence, or with a conventional case method, would help estimate the added value of DT-supported simulation. Delayed measurement would establish whether perceived effectiveness accompanies retention and transfer rather than only an immediate positive response.

Second, PEDTS should be expanded into a multi-item measurement model. Each domain should include several validated indicators, followed by confirmatory factor analysis, measurement invariance across disciplines or countries, and comparison of equal-weight, factor-score, and formative specifications. This program would determine whether PEDTS operates as a stable higher-order construct and whether the same meaning is preserved across student populations.

Third, future simulator studies should connect PEDTS to behavioral traces. Useful indicators include number and diversity of alternatives explored, frequency of iteration, response to feedback, time allocated to stakeholder analysis, and changes between initial and final decisions. Sequence analysis or process mining could then reveal whether high PEDTS emerges from deeper engagement with the DT cycle rather than from favorable post-hoc evaluation alone.

Fourth, sustainable decision quality should be assessed directly. A rubric could score environmental impact, social value, economic feasibility, systems thinking, stakeholder inclusion, and management of trade-offs. Multi-criteria decision analysis could compare the quality and balance of student choices across simulation rounds. This would allow researchers to test whether PEDTS predicts responsible decision improvement and whether that relationship is mediated by iteration or systems thinking.

Fifth, entrepreneurial development should be integrated through validated competency frameworks and artifact assessment. Opportunity recognition, value proposition coherence, experimentation, resource awareness, feasibility, and responsible value creation can be measured alongside the quality of prototypes or business models. A longitudinal design could examine whether DT-supported sustainability simulation strengthens entrepreneurial intention, self-efficacy, and capability without reducing sustainability to a secondary business constraint.

Sixth, the productive-challenge finding warrants a dedicated model. Future instruments should distinguish desirable difficulty, cognitive load, uncertainty tolerance, frustration, and exhaustion. A nonlinear or person-centered analysis could identify the point at which challenge supports learning and the point at which it becomes overload. This line of research could yield actionable design rules for calibrating simulation complexity and feedback.

Finally, the framework should be replicated through multi-institutional and cross-cultural studies. Multilevel modeling could separate student, team, instructor, course, and institutional influences. Measurement invariance and out-of-sample prediction would provide stronger evidence of generalizability. Open materials, preregistered hypotheses, shared code, and a common data dictionary would further support cumulative research and reproducibility.

VI. CONCLUSIONS

PEDTS provides a concise exploratory indicator of the perceived effectiveness of a DT-oriented sustainability simulation. Across 73 activity responses contributed by 66

students, internal convergence, the large PEDTS–UX association, and the ordered profile pattern remained substantively stable under participant-clustered inference, one-observation-per-participant sensitivity analysis, and permutation-based testing. The results support PEDTS positively as an evaluation layer while maintaining a clear distinction from full psychometric validation, objective sustainability competence, and entrepreneurial performance.

The main practical recommendation is to use PEDTS together with, rather than instead of, direct measures. PEDTS captures whether students perceived the intervention favorably as educationally valuable, meaningful, enjoyable, and engaging. Sustainability decision scores can establish whether choices became more responsible, while entrepreneurial rubrics can evaluate opportunity recognition and value creation. This combined assessment strategy would provide a richer account of how DT and simulation contribute to sustainable education.

Future studies should document each DT phase, measure decision changes across simulation rounds, separate productive challenge from overload, and compare PEDTS with sustainability knowledge, systems thinking, entrepreneurial competence, and solution quality. Stronger designs should add comparison groups, delayed measurement, behavioral logs, multi-item constructs, and multi-institutional replication. Such evidence can clarify not only whether students value a simulator, but which DT mechanisms generate that value and how the experience translates into responsible innovation, entrepreneurial action, and sustainable decision performance.

ACKNOWLEDGMENT

Identifying acknowledgments have been omitted to comply with the double-blind review process. They will be included in the accepted version of the manuscript

REFERENCES

[1] Green, C., Molloy, O., & Duggan, J. (2021). An empirical study of the impact of systems thinking and simulation on sustainability education. *Sustainability*, *14*(1), 394. DOI: 10.3390/su14010394.

[2] Novaes, André Lucas. "Enhancing sustainability education in higher education through simulation-based learning: integrating sustainable development goals." *International Journal of Sustainability in Higher Education* 27, no. 5 (2026): 1124-1141. DOI: 10.1108/IJSHE-08-2024-0571.

[3] Ardila Echeverry, M. P., Gauthier, A., Hartikainen, H., & Vasalou, A. (2025). Designing for Digital Education Futures: Design Thinking for Fostering Higher Education Students' Sustainability Competencies. *Sustainability*, *17*(10), 4289. DOI: 10.3390/su17104289.

[4] DT4S Project, "Design Thinking for Sustainability Education," 2022. [Online]. Available: https://dt4s.eu/.

[5] Baltador, L. A., Grecu, V., Panța, N. D., & Beju, A. M. (2024). Design thinking in education: Evaluating the impact on student entrepreneurship competencies. *Education Sciences*, *14*(12), 1311. DOI: 10.3390/educsci14121311.

[6] Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

[7] Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, *25*(2), 91-97.

[8] D. P. Ausubel, The Psychology of Meaningful Verbal Learning. New York, NY, USA: Grune & Stratton, 1963.

[9] ISO, B., & STANDARD, B. (2010). Ergonomics of human-system interaction. *British Standards Institution*.

[10] Sheffield, R., Koul, R., Bawa, S., Tee, L., Milbourn, B., & Ayoub, M. (Eds.). (2025). *Design Thinking for Sustainability Education: Utilising the Sustainable Development Goals for Impactful Teaching and Learning*. Taylor & Francis. DOI: 10.4324/9781003592129.

[11] Huang, et al. (2025). Designing for Digital Education Futures: Design Thinking for Sustainability. Sustainability. https://doi.org/10.3390/su17104289

[12] Vergara, D., Extremera, J., Rubio, M. P., & Dávila, L. P. (2019). Meaningful learning through virtual reality learning environments: A case study in materials engineering. *Applied Sciences*, *9*(21), 4625.