2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51418

Piloting a Brief, Cross-Disciplinary Survey Instrument for Measuring Sustainability Learning Outcomes: Year 2 Results from UTA's Engineering for One Planet (EOP) Institutionalization Program

Dr. Erick C. Jones Jr., The University of Texas at Arlington

Erick Jones is an assistant professor in the IMSE department at UTA and the founder and director of the Sustainable and Equitable Allocation of Resources or SEAR Lab. He obtained a PhD from the Operations Research and Industrial Engineering program at the University of Texas at Austin, a B.S. in Chemical Engineering from Texas A&M University, and is a fellow of GEM, NSF INFEWS, and DOE MLEF. He spent several years working in the design, manufacturing, oil and gas, and HVAC industries. During this time, he traveled around the world and witnessed how the lack of basic infrastructure like electricity, HVAC systems, clean water, internet, and banking negatively affects the quality of life of the majority of the world's population. These experiences motivated him to pursue research that can enhance quality of life by improving access to sustainable resources and economic opportunities, particularly where a lack of physical infrastructure or economic resources presents a major obstacle, leading to the creation of the SEAR lab. The SEAR lab investigates how communities, companies, and countries can allocate their limited resources in a way that maximizes their desired outcomes in a sustainable, equitable, and resilient but also elegant way. The SEAR lab assesses these problems by combining physical experimentation, data analytics, and stochastic systems optimization to provide actionable decisions and create scalable prototypes.

Mengqi Monica Zhan

MONICA ZHAN Dr. Zhan serves as an Associate Professor for the Department of Communication at the University of Texas at Arlington. Her research interests include employee communication and corporate communication. Recently, she has developed growing interests in underrepresented minority student and employee experiences.

GRACE BRANNON Dr. Brannon serves as an Associate Professor for the Department of Communication at the University of Texas at Arlington. Her research interests include communication pathways affecting specific outcomes (e.g., health, educational major choice) among marginalized populations.

LIWEI ZHANG Dr. Liwei Zhang serves as an Assistant Professor in the Department of Mechanical and Aerospace Engineering at the University of Texas at Arlington. Her research interests includes fluid and combustion physics, propulsion engineering, computational science, and engineering education.

FRANK LU Dr. Lu is a Professor in the Department of Mechanical and Aerospace Engineering at the University of Texas at Arlington. His research is in high-speed aerodynamics and propulsion, primarily in novel detonation engines.

Kendra Lee Wallis, The University of Texas at Arlington

Dr. Wallis did her academic work in physics and currently teaches in the College of Engineering at the University of Texas at Arlington. She is Director of Undergraduate Curriculum Development for the new Bachelor of Science degree program in Resource and Energy Engineering. She has been a Faculty Fellow for the Center for Teaching and Learning Excellence and currently leads the Professional Learning Community on Sustainability.

Ann M.L. Cavallo, The University of Texas at Arlington

Piloting a Brief, Cross-Disciplinary Survey Instrument for Measuring Sustainability Learning Outcomes: Year 2 Results from an Engineering for One Planet (EOP) Institutionalization Program

Abstract

This paper reports the design, administration, and preliminary results of a brief student survey instrument developed to measure sustainability learning outcomes in courses taught by faculty participating in the University of Texas at Arlington (UTA) Engineering for One Planet (EOP) Institutionalization Program. The program's original semester-long Sustainability Professional Learning Community (SPLC) proved difficult to scale during Year 1, and in Year 2 the team centered faculty development on a two-day Sustainability High Impact Practices (SHIP) bootcamp that was opened to faculty from any discipline. That pivot created a measurement problem: how to evaluate student-level impact in a large, heterogeneous set of courses without adding substantial burden on faculty or students.

To address this, we developed a 15-item survey organized into five constructs: Awareness and Exposure; Relevance and Learning; Knowledge and Definitions; Engagement and Behavior; and Overall Impressions. We administered the survey in Fall 2025 through SHIP-trained instructors in approximately ten courses spanning engineering, computing, and other disciplines. We obtained 246 responses from an undergraduate population that is 45.6% Asian, 25.2% Hispanic or Latino, 13.1% Black or African American, and 46.5% first-generation college students.

Across the sample, 80.3% of students reported that their instructor included sustainability content, 68.5% rated that content moderately or very relevant to their career, and over 70% reported that the material made the course more engaging. Goal 9 (Industry, Innovation and Infrastructure) was the most frequently recognized Sustainable Development Goal (SDG), with Goals 6 and 11 close behind. An initial large language model (LLM)-assisted thematic analysis of open-ended responses identified three recurring themes: real-world application, hands-on project work, and a broadening of students' definitions of sustainability beyond environmental concerns.

We report these findings alongside a set of item-level validity concerns the pilot surfaced, and we describe a planned validation pipeline that includes confirmatory factor analysis, human–LLM inter-rater reliability testing, and demographic disaggregation across cohorts. The results suggest that brief, course-embedded instruments can produce useful, comparable evidence of sustainability learning outcomes across disciplines, and they provide a foundation for the continued instrument development the program will undertake as it scales. An updated Spring 2026 dataset is being collected and will be incorporated in future work.

1. Introduction

Embedding sustainability across the undergraduate curriculum has become an increasingly common institutional priority, but the faculty development models used to enable that integration face persistent constraints: limited faculty time, uneven participation, and the difficulty of attributing student outcomes to brief pedagogical interventions [5, 16, 17]. At UTA, a Hispanic-Serving Institution (HSI) with a large first-generation undergraduate population, these constraints shaped our Engineering for One Planet (EOP) Institutionalization Program from the outset. In Year 1, a semester-long Sustainability Professional Learning Community (SPLC) drew only six faculty, while a two-day Sustainability High Impact Practices (SHIP) bootcamp attracted twenty-eight [31]. Faculty also told us, in interviews and informal feedback, that they wanted cross-disciplinary applicability and ready-to-use materials more than extended meetings. In Year 2 we therefore reorganized the program around the SHIP bootcamp and opened participation to faculty from any discipline.

That pivot created both an opportunity and a measurement problem. The opportunity was scale: a single bootcamp cohort now reached students across engineering, computing, mathematics, and beyond. The measurement problem was that the original evaluation plan, which relied on semester-long classroom observations and long-form faculty interviews, was no longer feasible for the number and variety of courses we now touched. We needed an evaluation instrument that was short enough to deploy in any course without substantial burden, meaningful across disciplines, and able to produce comparable quantitative indicators while still surfacing qualitative themes for program improvement.

This paper reports on the design and pilot administration of such an instrument. Specifically, we address two questions:

1. Can a brief, course-embedded student survey produce usable evidence of sustainability learning outcomes across a diverse set of courses and disciplines?
2. Based on the first semester of deployment, what does the instrument reveal about the student-level effects of the compressed SHIP workshop model, and what instrument refinements does it suggest for future rounds?

We report results from 246 students surveyed in Fall 2025 across approximately ten courses, together with an LLM-assisted thematic analysis of open-ended responses. We identify specific items that need revision and describe a planned validation pipeline for subsequent cohorts. We view this work as pilot evidence for a scalable, cross-disciplinary evaluation methodology that the engineering education community, and institutions considering EOP adoption in particular, can build on.

2. Background and Related Work

Five bodies of literature frame this study: high-impact educational practices, course-embedded assessment in engineering education, faculty development models for sustainability integration, service-learning and project-based approaches, and the use of the United Nations Sustainable Development Goals as a common language across disciplines. We draw on each briefly and then tie them together in a synthesis that motivates the specific instrument presented in this paper.

2.1 High-Impact Practices in (Engineering) Higher Education

Kuh's foundational work identified a set of High-Impact Educational Practices (HIPs), including learning communities, writing-intensive work, collaborative projects, undergraduate research, diversity and global learning, service-learning, internships, and capstones, as correlates of deep learning, retention, and student gains, particularly for historically underserved students [1, 2]. Subsequent syntheses and Association of American Colleges and Universities (AAC&U) implementations have reinforced the efficacy of HIPs and identified the design principles that make them high-impact: significant time and effort, interactions with faculty and peers, frequent feedback, reflective integration, authentic applications, and public demonstrations, with guidance to scale HIPs equitably across student populations [3, 4]. In engineering contexts, HIPs commonly appear as course-embedded projects, service-learning design experiences, and community-engaged capstones, which map well onto Accreditation Board for Engineering and Technology (ABET) outcomes and practice-based professional formation [21, 24]. Our own SHIP bootcamp takes its name from this literature: it is designed to equip faculty to introduce sustainability into existing courses through the HIP mechanisms most likely to produce deep, retained learning.

2.2 Course-Embedded Assessment and the Need for Brief, Valid Instruments

Engineering education's shift toward outcomes-based improvement has produced a steady call for valid, reliable, course-embedded measures that are sustainable for faculty and meaningful for students [5, 6, 7]. Systematic reviews find that rubrics and questionnaires dominate practice, but many measures lack documented validity and reliability, prompting calls for clearer construct definitions and professional assessment standards [5, 6]. ABET criteria institutionalize the continuous improvement cycle through direct and indirect assessment of student outcomes [8, 9, 10], reinforcing the demand for practical, course-level evidence that can be gathered without overwhelming instructors. Classic assessment work, including evidence-centered design and reasoning-from-evidence frameworks, argues for aligning claims, tasks, and observations in ways that are especially well suited to short, embedded instruments [7, 30].

For sustainability specifically, this demand is acute. The construct spans environmental, social, and economic dimensions, is typically introduced through short modules or embedded assignments rather than dedicated courses, and is delivered by faculty with widely varying backgrounds. A brief instrument designed to capture awareness, perceived relevance, and behavior or intent in any course that includes sustainability content therefore fills a documented gap. The instrument presented in this paper was developed with that gap in mind.

2.3 Faculty Development: Workshops, Bootcamps, and Communities of Practice

Evidence from science, technology, engineering, and mathematics (STEM) faculty development suggests that brief, focused workshops can shift instructional practice when they are paired with clear goals, active participation, and follow-up support [16, 17, 20]. Longer-format models, including faculty learning communities and communities of practice (CoPs), show stronger effects on sustained adoption but require substantial faculty time and often attract smaller, self-selected groups [17, 18, 19]. Our Year 1 experience, in which the two-day SHIP bootcamp attracted more than nine times the participation of the semester-long SPLC [31], is consistent with this pattern. Emerging models link CoPs, faculty mentoring networks, and the creation or adaptation of open educational resources to sustain reform and scale impact across institutions [18, 19].

An open empirical question is whether a compressed workshop, absent the extended community-of-practice scaffolding, can still produce measurable student-level effects. Short-form faculty development is common in part because it is feasible at scale. What is less well established is whether a single brief workshop is sufficient to register at the student level. This paper contributes preliminary evidence on that question.

2.4 Service-Learning and Project-Based Learning in Engineering

Engineering service-learning exemplifies HIPs that integrate real clients, societal needs, and iterative design. Studies report gains in critical thinking, intercultural competence, and professional outcomes when students participate for multiple semesters, with strong alignment to ABET outcomes [21, 22]. Systematic reviews of engineering service-learning assessment find widespread use of surveys and presentations, and call for more direct, outcomes-aligned measures, which reinforces the value of concise course-embedded instruments like the one we present here [23, 24]. UTA's own engineering faculty have an established track record of service-learning and community-engaged design, including research on sustainable engineering internships and multi-pronged curricular integration [31], and this capacity informs both the SHIP bootcamp curriculum and the research program in which this paper is situated.

Parallel to service-learning, project-based learning (PjBL) has accumulated a substantial evidence base. Recent meta-analyses conclude that PjBL produces statistically significant, positive effects on academic achievement, affective attitudes, and higher-order thinking skills, including within engineering and technology disciplines [25, 26, 29]. Reviews focused specifically on engineering report consistent benefits (engagement, real-world skills) alongside known challenges (workload, assessment complexity) and motivate the development of lightweight, embedded assessment tools that can capture outcomes without overburdening instructors or students [27, 28]. The combination of brief faculty preparation, such as our SHIP bootcamp, with course-embedded projects and concise evaluation is directly supported by this literature.

2.5 The UN Sustainable Development Goals as Interdisciplinary Framing

The SDGs have become a common reference point for sustainability integration in higher education because they work across disciplines and map onto familiar design constraints and learning targets [11, 12, 13, 32, 33]. Recent case work in civil engineering and across broader curricula shows phased approaches for bringing the SDGs into required sequences and aligning courses to SDG matrices, often with explicit attention to the social and economic pillars alongside environmental content [13, 14, 15]. In our SHIP bootcamp, the SDGs serve two functions. First, they provide shared language for faculty from disciplines that would not otherwise intersect. Second, they give students a framework for recognizing sustainability content when they encounter it, which makes them useful as survey items. The instrument described in this paper uses a single item asking students to select which SDGs they learned about during the course. We revisit that design choice in the Results.

2.6 Synthesis and Implications for This Study

Taken together, these literatures suggest three design principles for an evaluation instrument appropriate to our context. First, because HIP-aligned, course-embedded activities raise relevance and engagement when they are paired with valid, lightweight assessment, the instrument should be short enough to deploy without burdening faculty or students, and should target constructs (awareness, relevance, engagement, behavior or intent) that HIPs are known to move. Second, because SDG threading can be scaled through modular insertion into existing courses, the instrument should use the SDGs as an interdisciplinary common language while probing whether students recognize the framing. Third, because faculty development is most effective when short-form workshops are coupled with follow-up supports, the instrument should be sensitive enough to distinguish between courses in which implementation was stronger and weaker. The Year 2 redesign we implemented, combining a two-day SHIP bootcamp, cross-disciplinary reach, course-embedded sustainability activities, and a concise student survey, is consistent with the evidence base in each of these literatures, and it should be expected to produce measurable shifts in student awareness, perceived relevance, and engagement. The next sections describe how we operationalized these principles and what we found.

3. Program Context

The UTA EOP Institutionalization Program launched in Year 1 (2024–2025) with two faculty development pathways. The SPLC was a semester-long program with monthly meetings, intended as the primary mechanism for deep engagement with the EOP framework. The SHIP bootcamp was a two-day intensive, intended as an on-ramp to the SPLC. In practice, participation inverted: six faculty joined the SPLC while twenty-eight joined the SHIP bootcamp [31]. Faculty who attended both consistently told us that the bootcamp gave them enough to incorporate sustainability into one or more courses, while the SPLC served best as a deeper follow-on for a smaller subset of interested faculty.

Year 2 (2025–2026) restructured around this finding. The SHIP bootcamp became the primary entry point, and participation was opened to faculty from any discipline. Sustainability content was also integrated into other existing UTA professional learning communities, including ones focused on storytelling and on open educational resources, so we could reach faculty who would not self-select into a dedicated sustainability program. The Maverick Advantage course designation, which faculty can earn for courses that meet specific engagement criteria, continued to serve as a concrete incentive for incorporating sustainability content.

This paper reports only on the student-facing evaluation component of the program. We draw a deliberate distinction between the program, which creates training infrastructure, and the research reported here, which develops and tests measurement tools to assess that infrastructure's impact. The research uses the program as a research platform; the program would continue to operate without this research.

4. Methods

4.1 Instrument Design

We developed a 15-item survey organized into five constructs. The Awareness and Exposure construct includes items on prior familiarity with sustainability, whether the instructor included sustainability content, and the types of content encountered (select-all). The Relevance and Learning construct includes items on career relevance (Likert), agreement that the course increased disciplinary awareness (Likert), agreement that the course connected coursework to global issues (Likert), and the SDGs learned during the course (select-all). The Knowledge and Definitions construct includes a select-all definitions item and a true/false misconception check on whether sustainability is primarily environmental. The Engagement and Behavior construct includes items on course engagement, instructor encouragement to apply sustainability in future careers (Likert), and post-course behavior changes (select-all). The Overall Impressions construct includes an overall value rating and three open-ended questions. The instrument also collects demographic information (gender, race and ethnicity, veteran status, age, first-generation status) and administrative information (course and instructor).

Items were drafted by the research team, reviewed for face validity by two colleagues from the Center for Research on Teaching and Learning Excellence (CRTLE), and piloted informally with a small group of students before deployment. We did not attempt formal psychometric validation at this stage; the pilot is intended to produce data that will support such validation in subsequent rounds.

4.2 Administration and Sample

The survey was administered through QuestionPro under UTA Institutional Review Board (IRB)-approved protocols. SHIP-trained instructors distributed the survey link to students in courses where they had integrated sustainability content. Administration took place during Fall 2025. The landing page recorded 786 views and 287 survey starts. We used 246 submitted responses in the analyses below; denominators vary by item because some respondents did not answer every item, and all percentages reported here are computed from the number of respondents answering the item in question.

Respondents represented a demographically diverse undergraduate population. 36.2% identified as female and 61.8% as male, with 2.0% preferring not to say. Racial and ethnic identification was distributed across Asian (45.6%), White (26.3%), Hispanic or Latino (25.2%), Black or African American (13.1%), American Indian or Alaskan Native (3.5%), and Middle Eastern or North African (3.1%), with 8.5% preferring not to answer. 46.5% identified as first-generation college students; 84.3% were 25 years old or younger; 97.5% were not military veterans. This composition reflects the student population UTA serves and will allow us, in future analyses, to examine how responses vary across groups the engineering education field has identified as historically underserved.

4.3 Analysis Plan

The analysis plan has two components: a quantitative pipeline for closed-ended items and an LLM-assisted pipeline for open-ended items.

4.3.1 Quantitative Analysis.

For the current paper, we report frequency distributions for all categorical and Likert items, along with ranked bar charts for select-all items (types of content encountered, SDGs learned, definitions of sustainability, post-course behaviors). We do not yet report cross-tabulations by demographic group or discipline; both are reserved for confirmatory analysis after additional cohorts are collected. In subsequent rounds, we plan to conduct exploratory factor analysis to examine whether the five proposed constructs cohere as intended, to report internal consistency reliability

(Cronbach's α) for each subscale, and to conduct criterion-related validation against instructor-reported course integration intensity.

4.3.2 LLM-Assisted Qualitative Coding Pipeline.

The instrument includes three open-ended items. Manual coding at the scale this program will reach is not feasible, so we piloted an LLM-assisted thematic analysis approach in this round. The pipeline used for the current results has four steps:

1. Open-ended responses were exported from QuestionPro and stripped of identifying information.
2. Responses were submitted to a large language model using a structured prompt that asked the model to identify recurring themes, cluster responses under each theme, and return representative quotations.
3. Outputs were reviewed by two members of the research team, who read a random 20% sample of raw responses alongside the LLM-generated theme structure to verify that the themes surfaced genuine patterns rather than model artifacts.
4. The research team finalized the theme structure and drafted the narrative summaries reported in Section 5.6.

This pilot pipeline is intentionally light-touch. It uses the LLM for clustering and theme discovery while keeping interpretation in human hands. For subsequent cohorts we will formalize the validation in three ways. First, two trained human coders will independently code a baseline subset of responses ($n \geq 100$) using an inductively developed codebook, and inter-rater reliability will be established using Cohen's κ with a target of $\kappa \geq 0.70$. Second, the same subset will be processed through the LLM pipeline and human–LLM agreement will be reported using percent agreement and Cohen's κ; prompts will be refined iteratively until human–LLM agreement approaches human–human agreement. Third, we will conduct periodic audits of LLM output (10% sample) across subsequent semesters to monitor drift as the tool is used.

We view this methodology as one of the intended contributions of the broader research agenda that this paper introduces. Brief student instruments will increasingly be paired with open-ended items whose analysis is the bottleneck to deployment at scale, and a validated, auditable LLM-assisted pipeline would substantially reduce that bottleneck.

5. Results

Throughout this section we report results across all 246 respondents. We do not disaggregate by demographic group or discipline in the current round because per-cell counts would not support stable inference; disaggregation is a primary objective for the next round of data collection. Except where noted, percentages are computed from the number of respondents answering the item in question.

5.1 Demographics

The demographic composition reported in Section 4.2 indicates that the pilot reached a diverse sample. The preponderance of male respondents (61.8%) reflects the university's engineering student composition, while the 46.5% first-generation share and the racial and ethnic distribution reflect UTA's broader undergraduate population. This diversity, combined with a sample size approaching 250, is a methodological strength for the demographic disaggregation we plan in subsequent rounds.

5.2 Awareness and Exposure

Prior familiarity with sustainability varied before the courses began. 39.3% of students reported being moderately familiar, 32.6% slightly familiar, 15.5% very familiar, 2.9% extremely familiar, and 9.6% not at all familiar (Figure 1). This distribution confirms that UTA undergraduates generally arrive with limited prior exposure, which is consistent with the bootcamp's premise that faculty are reaching students who have not had a prior dedicated treatment of the topic.

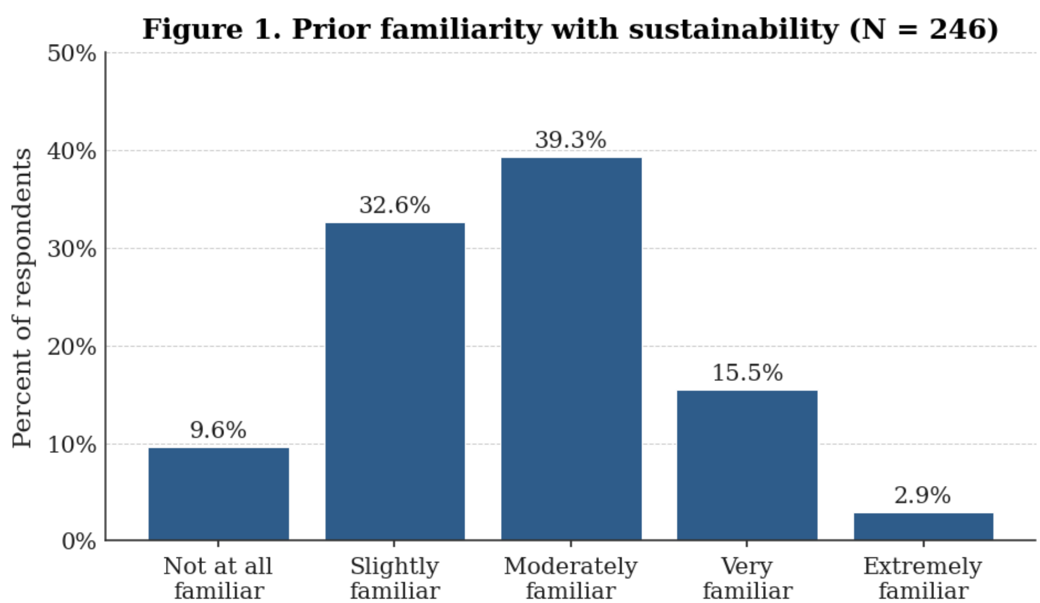

Figure 1. Prior familiarity with sustainability before the course (N = 246).

Of those surveyed, 80.3% reported that their instructor included sustainability-related content, examples, or activities in the course, 14.2% were unsure, and the remainder reported no such content (Figure 2). That roughly one in seven students in a course taught by a SHIP-trained instructor is unsure whether sustainability content was included bears closer attention in future rounds, both because it may indicate content embedded too implicitly to register with some students, and because it may identify courses where implementation was lower than intended.

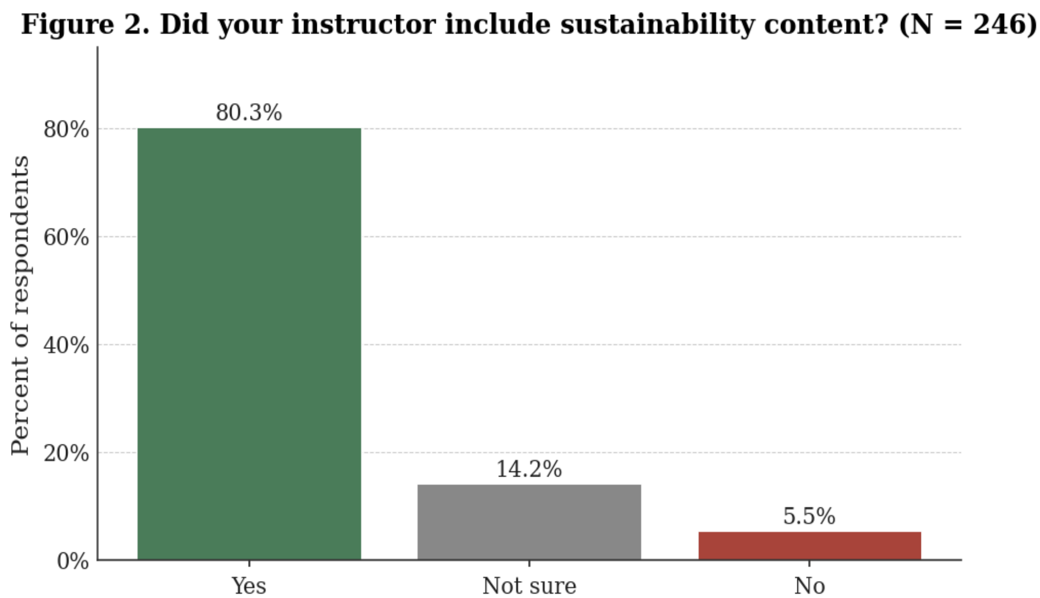

Figure 2. Student-reported inclusion of sustainability content in the course.

The types of content students encountered were relatively balanced across assignments or projects (23.8%), class discussions (22.2%), lectures or readings (22.1%), and examples linking sustainability to professional practice (19.9%), with explicit references to the SDGs less frequent at 11.8% (Figure 3). Given the role the SDGs play in the SHIP bootcamp itself, the gap between faculty framing and student recognition is a useful signal: students may be encountering SDG-aligned content without the SDG label attached. We return to this point in the Discussion.

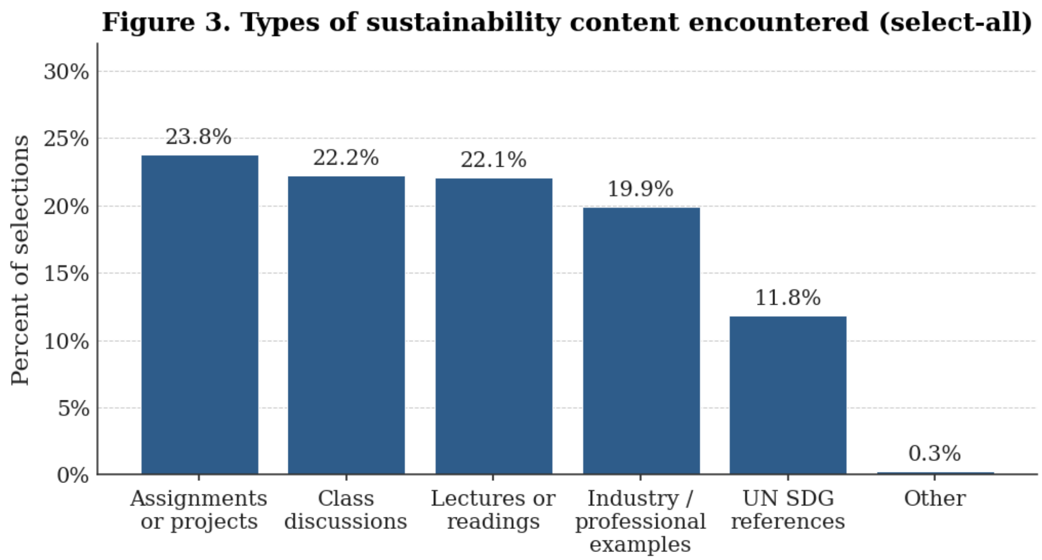

Figure 3. Types of sustainability content encountered (select-all).

5.3 Relevance and Learning

Students largely judged sustainability content to be relevant to their field of study or career. 38.4% rated it very relevant, 30.2% moderately relevant, and only five respondents rated it not relevant at all (Figure 4). In absolute terms, the proportion rating the material very relevant exceeded the proportion rating it moderately relevant, which was one of the more consistent findings in the pilot and one that held across the disciplines we could informally inspect.

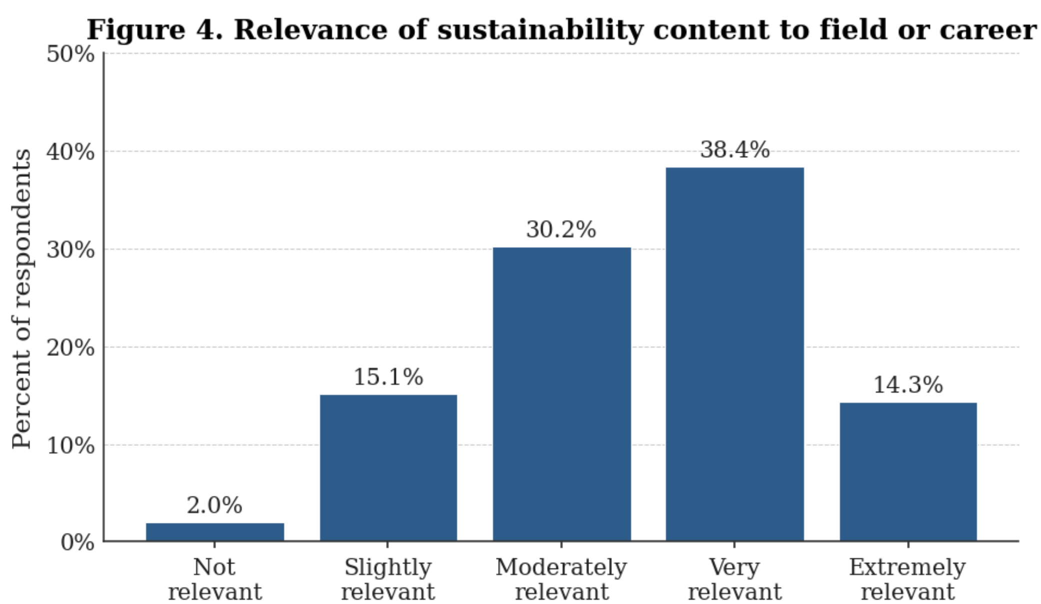

Figure 4. Perceived relevance of sustainability content to field or career.

69.1% of students agreed or strongly agreed that the course increased their awareness of how sustainability connects to their discipline, and 60.9% agreed or strongly agreed that it helped them see connections between global issues (such as poverty, climate change, and education) and their coursework (Figure 5). The difference between these two items is interpretable. The first asks about sustainability inside the discipline, while the second asks about connections from the discipline outward to global challenges. The slightly stronger endorsement of the first is consistent with faculty emphasizing applied, discipline-specific examples in their bootcamp-derived teaching.

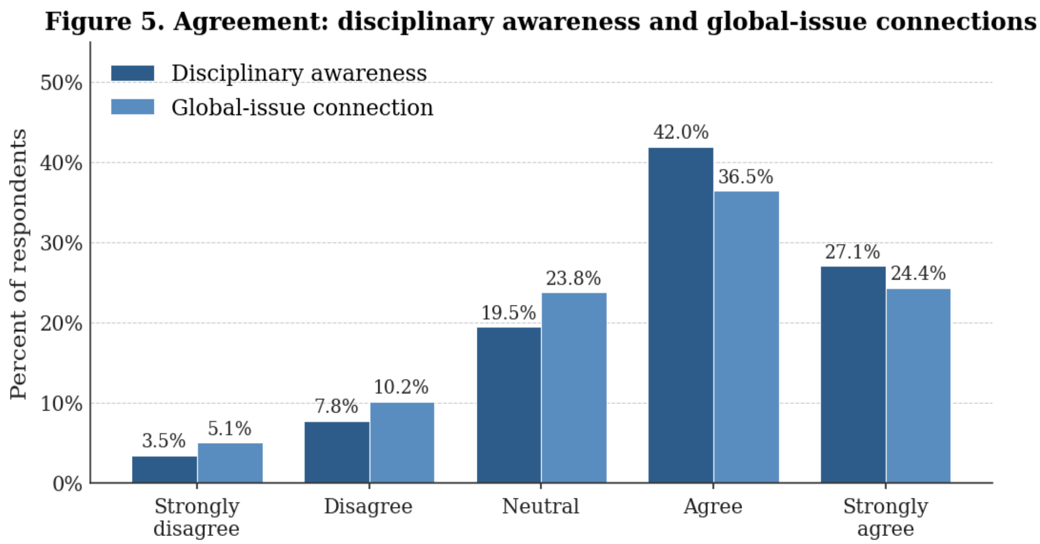

Figure 5. Agreement: course increased disciplinary awareness and global-issue connection.

On the SDGs learned item, Goal 9 (Industry, Innovation and Infrastructure) was the most frequently selected at 12.1%, followed by Goal 6 (Clean Water and Sanitation) and Goal 11 (Sustainable Cities and Communities), each at 8.5% (Figure 6). Goal 17 (Partnerships for the Goals) and several of the social-pillar goals were notably less frequent. We are cautious about interpretation here because the check-all format likely encouraged overselection in some cases, which is an item-level issue we discuss below.

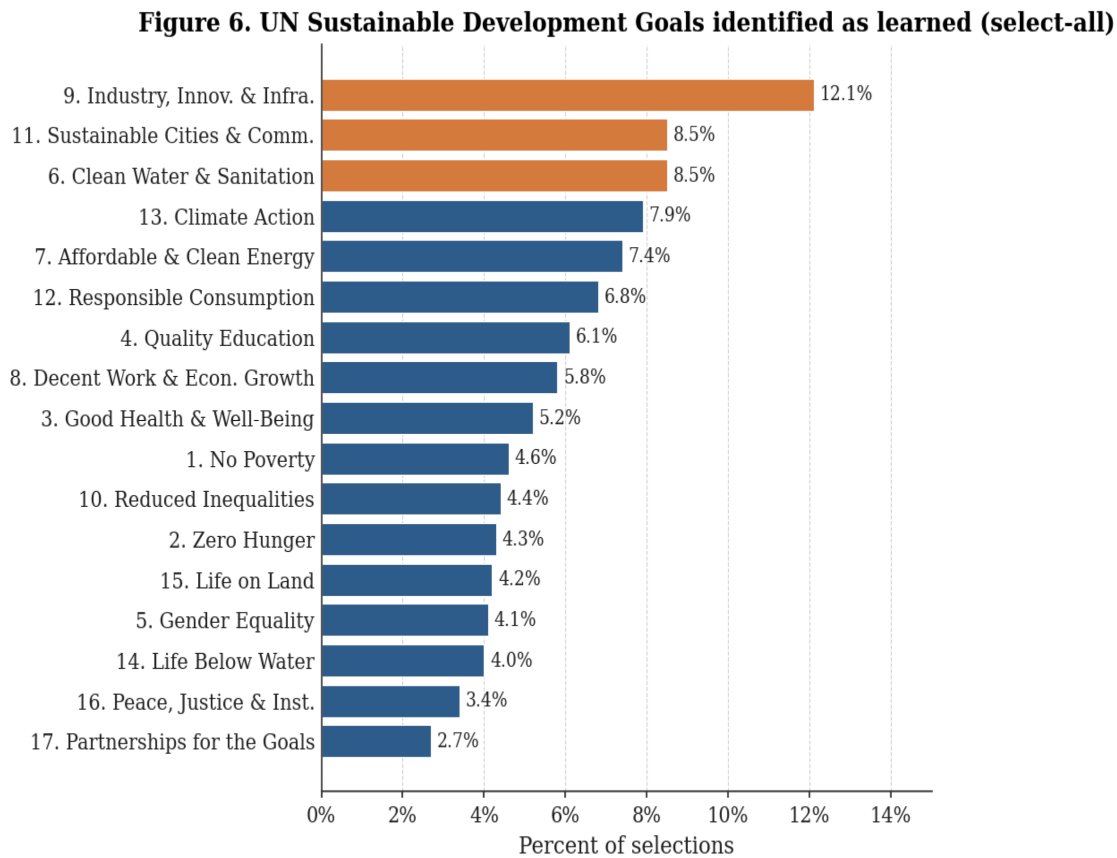

Figure 6. UN SDGs identified by students as learned during the course (select-all).

5.4 Knowledge and Misconceptions

Asked to select, from a list, all statements that correctly describe sustainability, students most frequently selected multi-dimensional definitions that include environmental, social, and economic considerations. However, in a separate true/false item designed to probe the common narrowing of sustainability to environmental issues alone, 56% of respondents endorsed the statement that sustainability is primarily environmental. On reflection, we consider this item a useful signal that a narrowing framing persists even after course exposure, but we also recognize that the statement as written admits a defensible reading under which a primary environmental emphasis is not strictly incorrect. Both items will be revised for Year 3.

5.5 Engagement, Behavior, and Perceived Value

Over 70% of students reported that the sustainability content made the course more engaging or interesting. 63.4% agreed or strongly agreed that their instructor encouraged them to think about how to apply sustainability in their future careers, and 85.1% rated the inclusion of sustainability content as moderately, very, or extremely valuable to their overall learning experience (Figure 7).

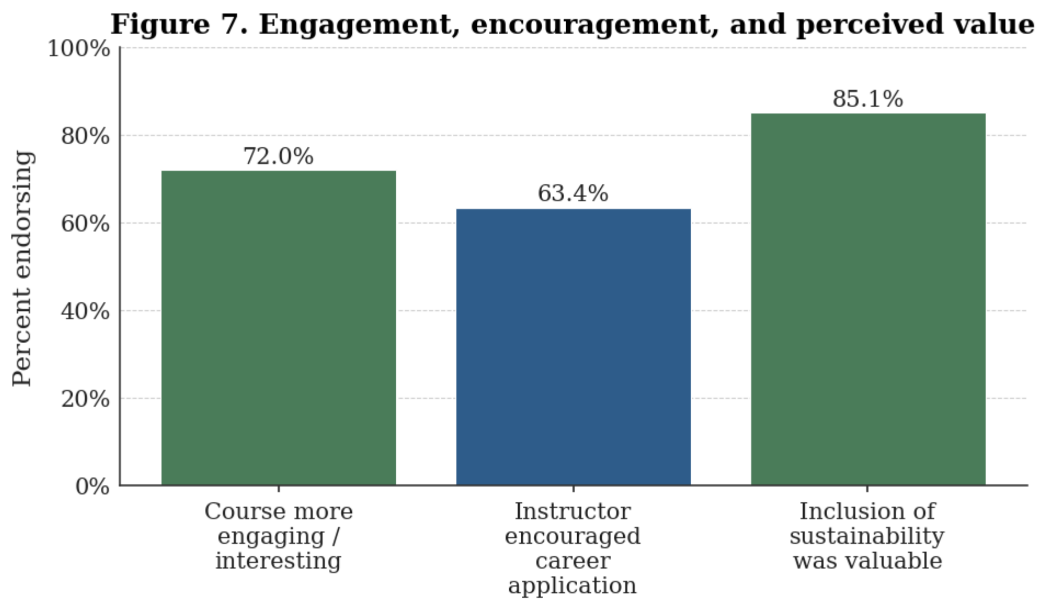

Figure 7. Engagement, instructor encouragement, and perceived overall value.

Self-reported post-course behaviors were mixed (Figure 8). 43% of students reported discussing sustainability topics with peers or family, 28% reported changing personal behaviors such as recycling or conserving energy, 26% reported considering sustainability-related career paths, and 13% reported that they had done nothing as a result of the course. We interpret the first three categories cautiously because social desirability may inflate self-reported behavior. Still, the pattern is consistent with other findings in the sample. Students most commonly translated the course content into further conversation, which is the lowest-cost form of engagement, while a smaller but meaningful fraction reported changes in behavior or career consideration.

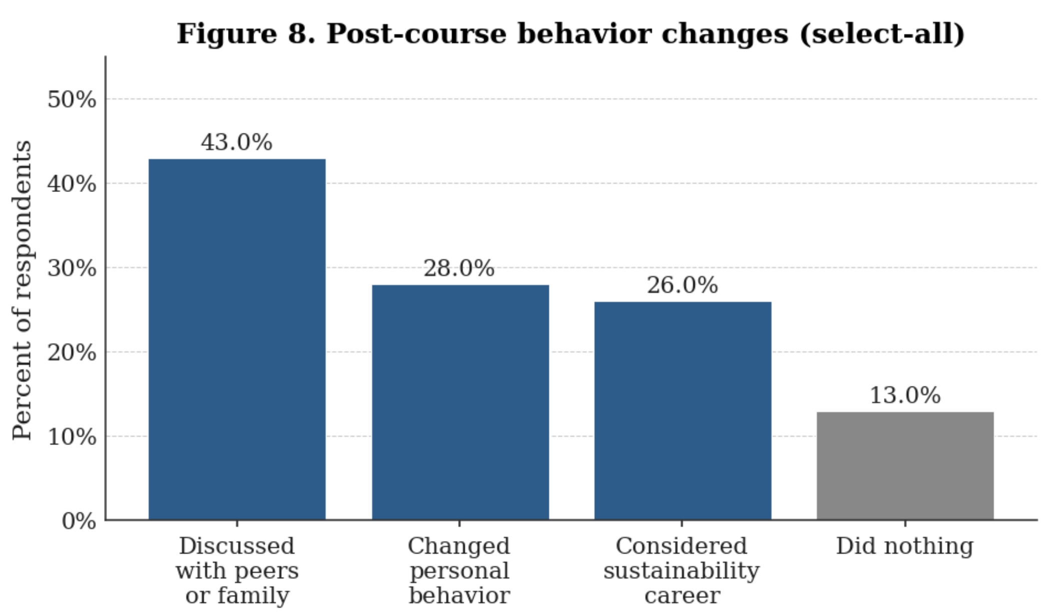

Figure 8. Post-course behavior changes reported by students (select-all).

5.6 Open-Ended Themes

The instrument included three open-ended questions. Using the LLM-assisted pipeline described in Section 4.3.2, we identified three main themes across the responses.

Real-world application drove engagement. Students most often described the sustainability content as engaging when it was tied to real problems. Responses referenced connections between engineering concepts and environmental or social challenges, the resonance of "real-world applications," and examples drawn from industry practice. A smaller set of students reported that sustainability was not consistently emphasized in their course, and this subset produced the most subdued engagement responses.

Hands-on work and the SDGs were the most memorable elements. When asked what was most memorable about learning about sustainability, students pointed to projects and assignments more than to lectures, and the SDGs appeared frequently as a framing device for those projects. Responses specifically referenced SDG-based group projects, water and wastewater treatment examples in engineering courses, and discussions of long-term consequences for communities and future generations.

Students' definitions of sustainability broadened during the course. Several responses explicitly described a shift from viewing sustainability as primarily environmental to seeing it as interconnected environmental, social, and economic systems. This theme, together with the quantitative misconception item, suggests that coursework is moving students toward a multi-dimensional view but that the narrowing frame is sticky and warrants more deliberate attention in future training materials.

Student suggestions for improvement, drawn from the third open-ended item, clustered around three requests: more real-world examples and case studies, more hands-on and applied work (mini-projects, simulations, site visits), and more consistent and frequent integration of sustainability throughout a course rather than as a single module. Taken together with the quantitative results, these suggestions point to directions for strengthening the SHIP bootcamp itself.

6. Discussion

We organize the discussion around three findings that we believe will be of use to others considering EOP adoption or designing similar evaluation systems.

6.1 A brief course-embedded instrument is workable across disciplines

The instrument administered in this study produced usable quantitative and qualitative evidence across a mixed set of engineering, computing, and non-engineering courses. The high share of students who recognized sustainability content in their course (80.3%), together with the distributional variation across items, indicates that the instrument is sensitive to differences in how instructors integrate the material. The LLM-assisted thematic analysis surfaced patterns that are plausible, pedagogically actionable, and triangulate with the quantitative findings. These are the basic conditions an instrument must satisfy to be useful at scale, and the pilot meets them. What the pilot does not yet establish, and what we turn to in Section 7, is formal psychometric validity across cohorts.

6.2 A compressed workshop model can shift student-level outcomes

Whether a two-day faculty workshop, absent longer-format community-of-practice scaffolding, can produce measurable student-level effects is an open question in the faculty development literature. Our pilot provides preliminary evidence that it can. Students in courses taught by SHIP-trained faculty consistently reported that sustainability content was relevant to their field (68.5% moderately or very relevant), that it increased their awareness of disciplinary connections (69.1% agree or strongly agree), and that it made their course more engaging (over 70%). These are indirect measures of faculty practice, but they are consistent with an implementation strong enough to register with students. For institutions facing the same resource constraints UTA faced in Year 1, the finding is pragmatic: a compressed, discipline-open workshop can be a defensible first-stage model while longer-format supports are built out.

6.3 Item-level findings point to concrete instrument revisions

Several items surfaced limitations that are themselves useful findings. The SDGs-learned item, presented as a check-all, likely elicited overselection and should be revised either to a ranking format or to a limited “select up to three” format. The misconception item asking whether sustainability is primarily environmental admits a plausible “true” reading and should be reformulated as a more specific contrast item. The engagement item is currently dichotomous and would benefit from conversion to a Likert format to capture variation. Each of these revisions is a straightforward instrument improvement, and each follows directly from this pilot’s data. We view the ability to diagnose and repair items in this way as an expected, desirable feature of iterative instrument development.

Beyond these item-level findings, the results have one broader implication. Students encountered SDG-aligned content at a higher rate than they recognized SDG framing (11.8% of exposure reports referenced the SDGs explicitly). If faculty are using the SDGs internally to structure their course design but not tagging that structure for students, students may benefit from more explicit framing. We will add this point to the SHIP bootcamp curriculum for Year 3.

7. Bridge to Future Research

The pilot reported here is the first stage of a multi-year instrument development effort. Our planned next steps, which together aim to move the instrument from pilot to validated, are as follows.

Psychometric validation. With the Year 3 cohort we will conduct exploratory factor analysis on the full item pool and examine internal consistency reliability for each of the five proposed constructs. We will then, with Year 4 data, conduct confirmatory factor analysis and report criterion-related validation against instructor-reported course integration intensity.

Longitudinal tracking. The current data represent a single semester. We will repeat administration each subsequent semester with refined items, which will allow us to estimate whether student-level effects persist beyond a single course and to examine cohort-level trends as the bootcamp evolves. A Spring 2026 dataset is currently being collected.

Demographic and disciplinary disaggregation. UTA's student population makes the instrument particularly well-suited to examining how sustainability content lands with students who are historically underrepresented in engineering. Disaggregated analyses by gender, race and ethnicity, first-generation status, and discipline are a primary objective of the next round.

Formal LLM validation. We will operationalize the validation protocol described in Section 4.3.2, with independent human coders establishing inter-rater reliability on a baseline subset, iterative prompt refinement until human–LLM agreement approaches human–human agreement, and periodic audits of LLM output as the pipeline is used across semesters.

Collectively, these activities constitute the research agenda of a separately proposed research project; the work described in the present paper is the pilot evidence motivating that agenda.

8. Limitations

The study has four main limitations. First, it is a single-semester, single-institution pilot and the findings are not intended to generalize beyond the courses represented. Second, item nonresponse varies across the instrument, which limits direct comparisons between items. Third, the current analysis does not disaggregate by demographic group or discipline, which means we cannot yet speak to differential effects. Fourth, the LLM-assisted coding pipeline in the current round relies on human sanity-checking rather than formal agreement statistics; the absence of that validation is one of our most important items of future work. We also note that students self-selected into the courses represented in the sample, and faculty are themselves self-selected into the SHIP bootcamp, which should be kept in mind when interpreting the positive direction of most findings.

9. Conclusion

A brief, 15-item student survey, administered across approximately ten courses taught by faculty trained through UTA's EOP Institutionalization Program, produced usable quantitative and qualitative evidence about how students experience sustainability content in their courses. The pilot results suggest that a compressed, two-day workshop model can register at the student level, that students connect sustainability content to their disciplines and careers, and that the SDGs function well as an interdisciplinary framing when faculty make the connection explicit. Item-level findings identify specific revisions for the next round, and a planned validation pipeline — including factor analysis, demographic disaggregation, and formal LLM–human agreement testing — will move the instrument from pilot to validated over the project's next phase. For institutions considering EOP adoption or similar sustainability integration efforts, the results offer both a workable evaluation approach and early evidence that such integration, even at modest intensity, can shift how undergraduates experience sustainability within their own fields.

References

[1] Kuh, G. D. (2008). High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter. Washington, DC: Association of American Colleges & Universities.

[2] Kuh, G. D., & O'Donnell, K. (2013). Ensuring Quality & Taking High-Impact Practices to Scale. Washington, DC: AAC&U.

[3] Finley, A., & McNair, T. (2013). Assessing Underserved Students' Engagement in High-Impact Practices. Washington, DC: AAC&U.

[4] Teaching & Learning Resource Center. (2023). High-Impact Practices: Enhancing the Student Experience. Ohio State University.

[5] Cruz, M. L., Saunders-Smits, G. N., & Groen, P. (2020). Evaluation of competency methods in engineering education: a systematic review. European Journal of Engineering Education, 45(5), 729–757.

[6] Douglas, K. A., & Purzer, Ş. (2016). Validity: Meaning and relevancy in assessment for engineering education research. Journal of Engineering Education, 105(1), 108–118.

[7] Pellegrino, J. W., DiBello, L. V., & Brophy, S. (2015). The Science and Design of Assessment in Engineering Education. In Cambridge Handbook of Engineering Education Research. Cambridge University Press.

[8] ABET. (2025). Criteria for Accrediting Engineering Programs, 2025–2026. Engineering Accreditation Commission.

[9] Kelly, W. E. (Ed.). (2008). Assessment in Engineering Programs: Evolving Best Practices. American Society for Engineering Education / AIR.

[10] Shryock, K. J., & Reed, H. L. (2021). ABET Accreditation—Best Practices for Assessment. ASEE Annual Conference & Exposition.

[11] Avelar, A. B. A., Oliveira, K. D. S., & Farina, M. C. (2023). The integration of the Sustainable Development Goals into curricula, research and partnerships in higher education. International Review of Education, 69, 299–325.

[12] Almssad, A., Almusaed, A., & Gbashi, G. Y. (2024). Engineering Curriculum SDG Integration: CDIO Standard 3 Approach. Proceedings of IHSES 2024.

[13] Hunashyala, A. M., Halyala, S., & Roopa, A. K. (2024). Path to SDG 2030: Fostering Sustainable Development at the School of Civil Engineering, KLE Tech—A Case Study. REES 2024 Proceedings.

[14] Bakthavatchaalam, V. (2024). Engineering Education and its Current (Un)suitability in Addressing Sustainable Development Goals. United Nations SDGs Knowledge Platform.

[15] University of Waterloo. (2024). Curriculum integration case studies—Sustainability in Curriculum. Institutional resource.

[16] Mataka, L. M., Saderholm, J. C., & Hodge, T. T. (2022). College STEM Faculty Teaching Practices: The Influence of a Professional Development Faculty Learning Community. Problems of Education in the 21st Century, 26(3), 40–56.

[17] Gehrke, S., & Kezar, A. (2017). The Roles of STEM Faculty Communities of Practice in Institutional and Departmental Reform in Higher Education. American Educational Research Journal, 54(5), 803–833.

[18] Kleinschmit, A. J., et al. (2023). Accelerating STEM education reform: linked communities of practice promote creation of OERs and sustainable professional development. International Journal of STEM Education, 10:16.

[19] Goldberg, B. B., Bruff, D. O., Greenler, R. M., et al. (2023). Preparing future STEM faculty through flexible teaching professional development. PLOS ONE, 18(10): e0276349.

[20] de Grave, W., Zanting, A., Mansvelder-Longayroux, D. D., & Molenaar, W. M. (2013). Workshops and Seminars: Enhancing Effectiveness. In Faculty Development in the Health Professions. Springer.

[21] Benning, J., Shearer, C., Kellogg, S., & Oakes, W. (2022). Impact of Service Learning on Engineering Student Development. International Journal of Engineering Education, 38(1), 253–263.

[22] Immekus, J. C., Maller, S., Tracy, S., & Oakes, W. C. (2010). Evaluating the Outcomes of a Service-Learning Based Course in an Engineering Education Program: Preliminary Results of the Assessment of EPICS. ASEE Annual Conference & Exposition.

[23] Queiruga-Dios, M., Santos Sánchez, M. J., Queiruga-Dios, M. Á., Acosta Castellanos, P. M., & Queiruga-Dios, A. (2021). Assessment Methods for Service-Learning Projects in Engineering in Higher Education: A Systematic Review. Frontiers in Psychology, 12:629231.

[24] Brown, A., & Bauer, M. (2020). Merging Engineering Education with Service-Learning. Journal of Pedagogical Research.

[25] Zhang, L., & Ma, Y. (2023). A study of the impact of project-based learning on student learning effects: a meta-analysis. Frontiers in Psychology, 14.

[26] Hidayat, H., Anwar, M., Harmanto, D., et al. (2024). Two Decades of Project-Based Learning in Engineering Education: A 21-Year Meta-Analysis. TEM Journal, 13(4), 3514–3525.

[27] Lavado-Anguera, S., Velasco-Quintana, P.-J., & Terrón-López, M.-J. (2024). Project-Based Learning (PBL) as an Experiential Pedagogical Methodology in Engineering Education. Education Sciences, 14(6), 617.

[28] Ferrari dos Santos, S. S. R., de Resende, L. M. M., & Xavier, J. C. (2024). Project-Based Learning in Engineering Education: A Systematic Review of the Literature. PAEE/ALE 2024 Conference Proceedings.

[29] Balemen, N., & Özer Keskin, M. (2018). The effectiveness of Project-Based Learning on science education: A meta-analysis. International Online Journal of Education and Teaching, 5(4), 849–865.

[30] Spurlin, J., Rajala, S. A., & Lavelle, J. P. (2023). Designing Better Engineering Education Through Assessment (2nd ed.). Taylor & Francis/CRC.

[31] Jones, E. C., Wallis, K. L., Zhan, M. M., Kim, N., Cavallo, A. M., & Clark, A. (2025, June). Teaching Teachers to Integrate Sustainability into Engineering Education: Lessons Learned from UTA's EOP Institutionalization Program. 2025 ASEE Annual Conference & Exposition, Montreal.

[32] Engineering for One Planet (EOP). (2024). The Engineering for One Planet Framework: Sustainability-Focused Learning Outcomes for Engineering Education. The Lemelson Foundation & VentureWell.

[33] Anderson, C., Hitt, S. J., & Truslove, J. (Eds.). (2024). The Engineering for One Planet Framework: Sustainability-Focused Learning Outcomes for Engineering Education with AHEP4 Mapping. The Lemelson Foundation.