5.3

How to Evaluate Education Outcomes

Ester Gimenez-Carbo[a], Adesola Adetunji[b], Franch Maverick A. Lorilla[c], and Irene Josa[d]

[a] Universitat Politècnica de València, Spain, ORCID: 0000-0002-2856-4081;
[b] University of North Dakota, USA, ORCID: 0009-0006-8067-8437;
[c] University of Science and Technology of the Philippines, the Philippines, ORCID: 0000-0001-6424-4600;
[d] University College London, United Kingdom, ORCID: 0000-0002-1538-4567.

Keywords: *programme evaluation, outcomes-based assessment, professional skills, rubric, student perception, evidence-informed practice*

Engineering programmes increasingly aim to develop graduates who can integrate disciplinary knowledge with transferable competencies such as communication, teamwork, ethical awareness and lifelong learning. To understand whether these outcomes are being achieved (and to improve teaching and curricula), educators need more than individual student assessments: they need systematic evaluation of outcomes at module, programme, and institutional levels. This chapter distinguishes assessment (methods that generate evidence of student performance) from evaluation (the process of interpreting aggregated evidence, together with stakeholder perspectives, to judge quality and inform decisions). Building on outcome-based education and quality assurance practices, we present a practical approach to designing an outcomes evaluation plan. The approach starts by clarifying purposes and evaluation questions, translating outcomes into indicators and performance criteria, and mapping where evidence is generated across the curriculum. It then combines direct and indirect measures to triangulate findings and strengthen interpretations. Finally, the chapter emphasises closing the loop: using results to guide curriculum development, educational policies, and staff development, and documenting continuous improvement for internal and external stakeholders.

DOI: 10.59490/mt.266.98.

1 Introduction

Engineering programmes have undergone a significant paradigm shift since the early 21st century, largely catalysed by the adoption of outcome-based accreditation standards such as the ABET Engineering Criteria 2000. While the integration of technical knowledge with broader professional skills has been a strategic goal for decades, these programmes increasingly prioritize the development of graduates who can seamlessly blend disciplinary expertise with transferable competencies — including communication, teamwork, ethical awareness, and critical thinking. In recent years, this approach has shifted towards managing complexity and sustainability, requiring engineers to possess not only technical skills but also advanced systems thinking to tackle interconnected global problems (Lönngren & Svanström, 2025). This intensified focus is driven by the complex demands of Industry 5.0 and the global urgency for sustainable development, which necessitate an unprecedented capacity for lifelong learning and ethical leadership (International Engineering Alliance, 2021; Muzulon et al., 2025; Sangwan et al., 2022).

However, declaring intended learning outcomes is not enough (Banta & Palomba, 2015). Programmes need credible evidence about whether students are developing these competencies and how learning opportunities across the curriculum contribute to that development (Crawley et al., 2014; Felder & Brent, 2003). Systematic evaluation of educational outcomes supports evidence-informed curriculum design and teaching enhancement, including checking that learning activities and assessment tasks are aligned with the intended outcomes (Biggs & Tang, 2011). It also provides transparency and accountability to stakeholders such as students, employers, professional bodies, and quality assurance agencies (Banta & Palomba, 2015).

In this chapter, 'assessment' refers to the methods used to generate evidence of learning (e.g., rubrics for projects, portfolios, peer and self-assessment, surveys, interviews, learning analytics). 'Evaluation' refers to the process of interpreting and triangulating that evidence (usually at cohort or programme level) to judge the extent to which outcomes are being achieved and to decide what should be improved. In other words, assessment produces evidence; evaluation turns evidence into decisions and action (Banta & Palomba, 2015; Olds et al., 2005). This distinction between assessment and evaluation reflects established strands in evaluation theory, particularly decision-oriented and use-focused approaches, which emphasise the role of evidence in informing judgement and improvement processes (Christie & Alkin, 2023).

Outcomes evaluation can take place at multiple, connected levels: (1) a teaching or curriculum intervention (e.g., introducing peer feedback to strengthen teamwork); (2) a module; (3) a full programme or curriculum; and (4) an institution or cross-institution initiative. Each level may require different evidence sources, time horizons, and stakeholders, but all benefit from a clear evaluation plan and an explicit 'closing the loop' process (Banta & Palomba, 2015; Felder & Brent, 2003).

Across Europe and beyond, outcomes-based approaches are increasingly embedded in internal quality assurance and accreditation processes (Augusti, 2009; Lucena et al., 2008). In this context, evaluation should not be treated as an administrative add-on: it is a mechanism for keeping skills education relevant, equitable, and sustainable over time.

This chapter supports educators and programme leaders by providing a practical framework. Section 2 explains how to use assessment evidence to evaluate programme outcomes and support continuous improvement. Section 3 explains how to design and run an outcomes evaluation plan that turns evidence into curriculum improvement, educational policy decisions, and staff development actions. Section 4 outlines approaches for evaluating specific transferable skills, and Section 5 provides illustrative examples and implementation tips.

2 Using assessment evidence for programme outcomes evaluation

This section is not about assessments and grading practices, but about the systematic use of assessment data to evaluate educational outcomes and improve programmes. Assessment determines whether individual students have met the defined criteria for module passing. In contrast, evaluation interprets aggregated assessment evidence over time to determine the extent to which programme-level education outcomes are being achieved and to inform module contents, curriculum development, quality enhancement, and policy decisions.

2.1 Types of evidence for programme outcomes evaluation

Evaluating educational outcomes requires drawing on multiple complementary sources of evidence, as no single measure is sufficient to determine whether intended outcomes, such as transferable skills and competencies, are being achieved consistently across a cohort. Effective evaluation, therefore, depends on the systematic collection, aggregation, and interpretation of structured data. There are basically two types of evidence required for evaluating education outcomes: direct evidence and indirect evidence — which often generate a combination of quantitative and qualitative data.

Direct evidence of outcome attainment. Student work that is clearly aligned with specific educational outcomes serves as direct evidence. Examples of such evidence include the writing and presentation of final-year dissertations, capstone projects, laboratory assignments, design reports, and weekly project progress presentations. When these artefacts are evaluated against appropriate performance rubrics or criteria, they provide measurable indicators (often in the form of quantitative data) of whether the educational outcomes and module objectives are achieved. The importance of these artefacts for assessing students lies not in individual grades but in overall performance trends. For instance, rubric data collected from several students over time can highlight consistent weaknesses in critical thinking or problem solving, as well as recurring strengths in teamwork or communication. These insights can inform curriculum improvements or changes in teaching methods. The rationale for using different evidence measures is their proximity to authentic performance. Direct evidence artefacts and data enable programme teams to evaluate whether students can demonstrate the intended transferable skills and competencies in realistic contexts. In this context, Sandoval (2014) argues that participant artifacts are essential because they serve as 'proxies' for learning processes. These artifacts indicate the extent to which learners are engaged in the specific activity and thinking hypothesized to be significant for achieving the desired outcomes.

Indirect evidence and perception-based measures. Surveys and feedback from students, alumni, instructors, and employers serve as indirect indicators of educational quality, providing insight into learners' perceptions of their growth and readiness. These artefacts, while they do not directly demonstrate competencies and transferable skills, offer valuable qualitative insights into how programme education outcomes are perceived, experienced, and valued by different stakeholders. At the programme level, these artefacts also serve important evaluative functions, as they can reveal perceived gaps in the curriculum and teaching approaches, track students' development and confidence in transferable skills, and enable comparisons between internal performance patterns and external expectations. It is important to note that indirect evidence must be interpreted with care, as response bias, cultural differences, and varying stakeholder expectations can influence it. For this reason, indirect indicators should supplement, rather than replace, direct evidence of student learning and performance. The rationale for integrating indirect evidence is that programme quality encompasses not only demonstrable performance at the programme level, but also stakeholder perceptions of satisfaction, the degree to which the curriculum supports employability, and its alignment with contemporary professional practice. According to recent studies, perception surveys provide more information when interpreted alongside direct evidence of learning. Without this cross-referencing of data, individual instruments often do not offer sufficient validity for programme evaluation (Osipovskaya et al., 2024).

The distinction between assessment and evaluation rests on the aggregation and interpretation of evidence. Module-level data must be systematically aligned with programme-level outcomes through processes such as consolidating rubric scores, examining longitudinal trends, comparing cohorts, and identifying acceptable attainment thresholds. Analytical tools support this work by revealing patterns that inform evidence-based judgement. Triangulating quantitative and qualitative data from direct, indirect, and institutional data strengthens the credibility of conclusions, while structured analysis ensures decisions are grounded in coherent evidence rather than isolated impressions. To put this distinction into practice, institutions use frameworks based on defined indicators and strategic assessments. This provides reliable and scalable evidence of student performance at programme level (Mahalingam, 2026).

Furthermore, evidence (both direct and indirect) on educational outcomes is valuable only when it is part of a clear evaluation framework in which data are analysed against predefined criteria and directly inform programme decisions. This systematic approach aligns with the technique of 'conjecture mapping', which Sandoval (2014) defines as a means of specifying theoretically salient features of a learning environment, mapping how they are predicted to work together to produce desired outcomes. Such a framework provides an 'argumentative grammar', the logic that guides a method and supports reasoning about its data, clarifying how practical improvement and theoretical refinement occur concurrently. This includes assessing whether desired outcomes are met, identifying gaps or overlaps in the curriculum, guiding targeted actions like curriculum redesign or staff training, and tracking the results of these actions. In this broader evaluation process, assessment tools serve not only as methods for measuring learning but also as sources of practical insights that help to continuously improve the program.

Integrating direct and indirect data and analysing them within a clear framework provides a coherent understanding of educational quality and informs curricular decisions that support continuous improvement and alignment with evolving professional demands. This systematic classification of evidence ensures that the evaluation is grounded in a logical structure, rather than isolated observations, providing what is known as an argumentative grammar for the assessment process.

The identification and analysis of the different sources of evidence provide the foundation for determining what information can credibly support judgements about the achievement of learning outcomes. However, having evidence is not enough on its own: a structured process is required to decide what will be evaluated, how the evidence will be interpreted, and how conclusions will translate into concrete improvement actions. Section 3 addresses this need by presenting a practical framework for designing and implementing an outcomes evaluation plan that turns collected evidence into coherent decisions at module, programme, or institutional level.

3 Practical solutions for an effective outcomes evaluation plan

Once transferable skills and other intended learning outcomes are embedded in modules and curricula, educators and programme leaders need a structured way to determine whether those outcomes are being achieved, for whom, and under what conditions. An outcomes evaluation plan describes (1) what will be evaluated; (2) which evidence will be used; (3) how findings will be interpreted; and (4) how improvement actions will be agreed upon, implemented, documented, and reviewed.

3.1 Principles for a credible and useful outcomes evaluation

Effective outcomes evaluation is 'fit for purpose', that is, it is designed around decisions that need to be made, collects only evidence that will be used, and makes explicit how evidence will be turned into improvements. The following principles help ensure that conclusions are credible and that the process is sustainable.

Start with identifying intended users and decisions to be made. Clarify who will use the findings (e.g., teaching teams, programme committees, students, employers, quality assurance or accreditation reviewers) and what decisions the evaluation needs to inform (e.g., curriculum redesign, staff development priorities, resourcing).

- *Prioritise validity of inferences.* Use evidence that genuinely represents the intended outcomes (especially complex transferable skills), ensuring that the evidence maps directly to the intended construct (e.g., content-based evidence) and avoid over-interpreting single tasks or single cohorts.
- *Support consistency and reliability.* Where human judgement is involved (e.g., rubric scoring), use assessor calibration, moderation, and stable scoring guidance so that trends across cohorts can be interpreted with confidence.
- *Build in triangulation.* Combine multiple sources and perspectives (direct and indirect, quantitative and qualitative) so that conclusions do not rely on a single measure.
- *Attend to fairness and equity.* Check whether evidence generation or interpretation disadvantages particular student groups; where possible, review patterns using disaggregated data and consider contextual explanations before acting.
- *Be transparent and proportionate.* Make outcomes, criteria, data sources, and limitations visible to stakeholders, and design a process that is feasible within available time and resources (e.g., sampling, rotating outcomes on a multi-year cycle, reusing existing assessment data).

3.2 Building the evaluation plan: a practical workflow

The workflow below can be adapted to different contexts (single module, full programme, or institution-wide initiative). It is deliberately framed around outcomes evaluation rather than the design of individual assessment tasks.

1. *Define the scope and evaluation questions.* Specify what level is being evaluated (module, programme, institution), which outcomes are in scope, and what decisions the evaluation should enable. Keep the initial set of questions small (e.g., two to four questions) and relevant to the decisions to be made.
2. *Translate outcomes into indicators and success criteria.* For each outcome, define a small number of observable indicators and set an acceptable achievement (performance criteria or targets). For transferable skills, indicators often need to be context-specific (e.g., teamwork in design projects versus teamwork in laboratories).
3. *Map where evidence is generated.* Use a curriculum map (or evidence map) to identify where each outcome is introduced, developed, and demonstrated, and to select a few high-value evidence points (often called signature assessments) that will be used for evaluation.
4. *Select evidence sources and a sampling strategy*. For each outcome, choose two or three complementary sources (e.g., rubric-rated performance in authentic tasks plus a student survey and/or supervisor feedback). Decide which cohorts will be included, how often evidence will be collected, and whether sampling is needed to keep the workload manageable.
5. *Plan data management and analysis.* Assign responsibilities (who collects, anonymises, analyses), specify timelines, and define how evidence will be stored securely. Predefine simple analysis approaches (e.g., cohort-level

distributions, trends over time, subgroup patterns, thematic analysis of qualitative comments).

6. *Aggregate, triangulate, and interpret.* Bring evidence together at the cohort or programme level and check for convergence or plausible reasons for divergence across sources. Interpretation should include uncertainty: what the evidence supports strongly, and what requires caution or further data.
7. *Decide, act, and 'close the loop'.* Hold a structured evaluation (sense making) meeting to agree conclusions and actions. Translate conclusions into a time-bound improvement plan (curriculum changes, assessment policy adjustments, learning support, staff development), assign owners, and document what was changed and why (Felder & Brent, 2003).
8. *Review impact and improve the evaluation process itself.* In the next cycle, check whether actions corresponded to improvements in the targeted outcomes, and refine indicators, evidence points, or analysis methods where needed (meta-evaluation) (Banta & Palomba, 2015).

3.3 Minimum documentation to support continuous improvement

To keep outcomes evaluation manageable, it helps to maintain a small set of living documents that can be updated each cycle. A typical minimum set includes:

- An outcomes evidence map (matrix) showing which modules and activities provide evaluative evidence for each outcome;
- Definitions of indicators, performance criteria or targets, and any rubrics used for programme-level judgements;
- An evidence calendar (what is collected, from whom, and when);
- An annual or cyclical evaluation summary that records the evidence reviewed, conclusions reached, limitations, and agreed actions;
- An improvement log tracking actions, owners, timelines, and follow-up evidence ('closing the loop' documentation).

Once an evaluation plan is established and a systematic process for turning evidence into curricular and organisational improvements is in place, a further challenge arises: evaluating the transferable competencies that are essential in engineering education. These competencies — such as problem solving, critical thinking, design, teamwork, and communication — require tailored methods and context-appropriate approaches due to their complexity.

4 Examples

The following examples illustrate how the principles outlined in the previous sections can be applied in real educational contexts. Each example highlights different dimensions of outcomes evaluation, such as the use of authentic assessment evidence, triangulation across data sources, faculty calibration, or the integration of evaluation results into programme-level decisions. Together, these examples demonstrate how structured evaluation processes operate in diverse institutional settings and provide practical insights that reinforce and complement the conceptual framework developed in Sections 2 and 3. Examples 1 and 2 operate primarily at the module level, demonstrating how systematic assessment of transferable skills generates the evidence base that feeds into programme-level quality assurance cycles and improvement decisions. Example 3 illustrates evaluation at the full programme level. For more information on how to teach and assess these competencies, we refer you to the relevant chapters in Part 2 and 3 of this handbook.

4.1 Evaluation of outcome-based assessment through project-based learning in electronics engineering at USTP

The University of Science and Technology of Southern Philippines (USTP) is a public state university located in Mindanao, Philippines, offering accredited engineering programmes aligned with national and regional quality assurance frameworks. Like many institutions operating within outcome-based education systems, USTP has progressively integrated programme-level evaluation mechanisms to support continuous quality improvement and international benchmarking.

At the USTP, the College of Engineering implemented a refined outcomes-based education (OBE) framework through a project-based learning (PBL) approach in its Electronics Engineering capstone design module. This initiative aimed to enhance the assessment of both technical and transferable skills such as communication, problem solving, and teamwork.

Context and implementation

The module required student teams to design, prototype, and test Internet of Things (IoT) solutions addressing local community needs (e.g., water-level detection for flood-prone areas, smart energy meters for small enterprises). Projects were aligned with ABET and the Philippines Commission on Higher Education (CHED) engineering outcomes, with particular focus on, problem solving, design and experimentation, communication, ethics, and lifelong learning

Table 5.3.1: Assessment tools and techniques used

Tool/technique	*Purpose*	*Frequency*	*Platform*
Rubrics (holistic and analytic)	Evaluate design quality, innovation, documentation, and teamwork	Per milestone	Google Docs/Canvas LMS
Peer and self-evaluation	Reflect on team contributions and learning	Biweekly	Microsoft Forms
Faculty feedback loop	Weekly design check-ins and troubleshooting	Weekly	In person and online
Formative surveys	Capture ongoing learning challenges	Midterm and finals	Microsoft Forms
Capstone pitch scoring	Real-time evaluation during Demo Day	End of module	Panel (academic and industry)

Rubrics were structured on a four-level scale (Beginning, Developing, Proficient, Advanced), with weighted criteria assigned to technical accuracy (40%), design innovation (25%), communication clarity (20%), and teamwork and process documentation (15%). Faculty calibration sessions were conducted prior to milestone grading to improve inter-rater reliability. Peer assessments contributed 10% of the teamwork component to promote accountability.

Quantitative outcomes

- 80% of teams improved their design scores by at least one rubric level (e.g., from Developing to Proficient) between milestone 1 and Demo Day.
- 94% of students rated the feedback and peer assessment system as 'clear' or 'very clear', based on end-of-semester surveys (n=72).
- 75% of students identified teamwork as their most improved skill through reflective self-assessments.
- Faculty scoring variance reduced by 28% over two terms after rubric calibration and peer norming.

What worked

- Scaffolded assessments, in which rubrics were layered across phases (ideation → prototyping → deployment) helped reinforce key outcomes without overwhelming students.
- Peer and self-assessment empowered students to take ownership and reflect on their roles, improving engagement and group accountability.
- Feedback loops with weekly consultations and digital feedback tools shortened the feedback cycle, allowing for immediate iteration.

What we learned

- Rubric clarity is key. Language must be aligned with student comprehension levels; visual anchors (e.g., sample outputs) helped a lot.
- Tracking progression matters. Comparing early-stage and final outputs gave students a tangible sense of growth.
- Team dynamics must be managed. Some high-performing groups still experienced internal conflict; proactive monitoring helped address this early.

Institutional alignment

This module design aligned with USTP's academic KPIs and supported programme-level quality assurance processes. In the Philippine context, this included compliance with regional accreditation frameworks such as AUN-QA, the ASEAN University Network Quality Assurance (ASEAN University Network, n.d.) and national accreditation bodies. Comparable processes exist within the European Higher Education Area under the Bologna Process and ESG quality standards. The aggregated rubric data, peer assessment records, and survey results from this capstone module serve as direct and indirect evidence inputs into USTP's annual programme evaluation cycle, informing decisions on curriculum sequencing, faculty development priorities, and learning support provision at the programme level.

4.2 Evaluation of Simulation-supported reflective assessment in process engineering education

At the University of North Dakota, a graduate-level chemical engineering module integrated reflective practice with Aspen Plus simulations to evaluate students' communication and critical thinking skills. As part of a midterm assignment, students designed a sustainable ammonia production process using green hydrogen. The project required them to consider environmental impacts, energy balances, and cost efficiency, bridging engineering technical skills with sustainable systems thinking.

Quantitative outcomes

- 85% of students showed measurable improvement in their communication skills and ability to justify engineering decisions based on rubrics.
- 78% of students demonstrated greater self-awareness and critical reflection in their journal entries.
- Peer evaluations highlighted and helped address communication gaps within student teams.

What worked

- Combining simulation tasks with reflective journaling fostered deeper learning and higher-order thinking skills.
- Structured feedback loops, both formative and summative, supported continuous improvement throughout the semester.
- Including multimodal communication (written reports, presentations, journals) engaged diverse learners and built confidence in expressing technical ideas.

Table 5.3.2: Assessment tools and techniques used

Tool/technique	*Purpose*	*Frequency*	
Simulation reports	Assess technical understanding and optimisation decisions	Midterm and final	Aspen Plus, PowerPoint
Reflective journals	Capture reasoning behind choices and group dynamics	Biweekly	Google Docs
Oral presentations	Demonstrate and defend design choices	End of module	In-person and online
Peer evaluation	Assess communication and contribution to teamwork	End of module	In-person and online

What we learned

- Providing clear rubrics and explicit guidance early is essential for supporting students in reflective and communication tasks.
- Students benefit from early exposure to professional expectations for technical communication and justification of design choices.

- Incorporating analytical tools such as simulations alongside reflective practice strengthens transferable skills for professional engineering work.

Institutional alignment

This practice supports undergraduate and graduate programme-level outcomes on critical thinking and communication. Lessons from this case have been included in departmental quality assurance reviews, leading to curriculum updates with a stronger focus on communication, reflective practice, and industry relevance. The triangulated evidence generated through simulation reports, reflective journals, oral presentations, and peer evaluations has been used as direct and indirect input for programme-level review meetings, helping faculty identify where communication and critical thinking outcomes need stronger scaffolding across the degree sequence.

4.3 A programme-level outcomes evaluation approach: EDINSOST (Spain)

Sustainability is increasingly treated as a transversal engineering competency, closely connected to professional judgement, ethics, systems thinking, and design decision making. Because sustainability is developed across multiple subjects (rather than in a single module), it is particularly well suited to programme-level outcomes evaluation: the goal is not to score individual students, but to determine where, how, and how consistently sustainability-related learning outcomes are embedded in the curriculum, and whether students perceive that they are actually achieving them.

The EDINSOST initiative, a cross-university Spanish research project named "*Education and Social Innovation for Sustainability"*, provides a structured and operational model for this purpose. It translates sustainability into an explicit Sustainability Competency Map (SCM), in which the competence is decomposed into competency units and specific learning outcomes, organised across progressive domain levels. This mapping offers a clear reference of expected graduate capabilities and serves as a backbone for both curriculum alignment and evaluation (Sanchez-Carracedo et al., 2021b, 2022). The learning outcomes in the SCM are classified using a three-level domain taxonomy based on a simplified Miller's pyramid, i.e., a framework created by George Miller in 1990 that outlines progressive levels of professional capability, commonly used in medical and healthcare education,(Miller, 1990): Knowing (level 1), which refers to theoretical learning; Knowing How (level 2), which involves integration and development in specific situations; and Demonstrating + Doing (level 3), which is linked to showing the competency in action and the possibility of transferring it to professional practice. This structured hierarchy allows the SCM to be easily embedded into the degree curriculum by assigning specific domain levels to different subjects along the competency itinerary.

From an outcomes evaluation perspective, EDINSOST combines multiple complementary data collection and analyses processes to construct a robust programme-level picture:

Curriculum 'presence' mapping (intended/enacted curriculum). a presence map is produced by identifying which subjects develop which sustainability learning outcomes. The approach can use documentary analysis (e.g., subject guides) and/or faculty input. To ensure practical reliability, the construction of these presence maps follows a two-step methodology: first, faculty members complete a diagnostic survey; second, experts in education for sustainable development (ESD) conduct semi-structured interviews with those professors. This expert-led intervention is critical to unify reporting criteria across different departments and to mitigate the Dunning-Kruger-effect, where instructors less familiar with ESD may inadvertently overestimate the depth to which sustainability is addressed in their classroom activities.

External referencing to the SDGs (accountability and coherence). EDINSOST also illustrates how to connect programme outcomes to wider societal frameworks. A correspondence matrix can be developed to map programme sustainability learning outcomes to the SDGs (and associated learning objectives), helping programme teams see which SDGs are already supported, where adaptation is needed for engineering contexts, and where gaps remain.

Triangulation with student evidence (experienced/learned curriculum). to evaluate whether the curriculum embedding translates into student learning, EDINSOST proposes using a student survey to capture students' perceived achievement of the sustainability learning outcomes. This assessment of the 'learned curriculum' is conducted through a validated 34-item Engineering Sustainability Questionnaire (ESQ). Students respond to statements using a four-point Likert scale, which enables a pre-post analysis. By comparing the perceived learning of first-year students against those completing their bachelor thesis, programme teams can quantify the relative learning increase across each of four sustainability competencies, holistic, environmental, social, and economic (Sanchez-Carracedo et al., 2021a, 2021c).

This approach supports continuous improvement because the outputs are directly actionable: they identify strengths, weaknesses, and inconsistencies across the programme and across institutions. For example, EDINSOST-related analysis has been used to examine how sustainability is unevenly present across degree programmes and universities, demonstrating how an outcomes evaluation lens can surface curricular disparities that are not visible from single-module assessments alone.

Comparing student perceptions with the presence map can reveal 'alignment gaps' (e.g., outcomes that staff believe are covered but students do not recognise as learned). These findings are directly actionable: they guide the design of targeted faculty training courses and the collection of open educational resources to help teachers embed specific missing learning outcomes into their subjects. Thus, the evaluation plan ceases to be a mere administrative report and becomes a catalyst for curriculum redesign.

As a practical note, because presence maps primarily reflect the intended and enacted curriculum (according to documents and staff reporting) and perception surveys capture indirect evidence, programme teams can strengthen conclusions further by adding direct evidence from selected assessment artefacts (e.g., capstone design rubrics or project portfolios) aggregated at cohort level, especially when high-stakes 'closing the loop' decisions are required.

5 Final remarks

Outcomes evaluation is more than a technical process — it is a mechanism through which institutions learn about themselves. Across the cases and methods described in this chapter, a common message emerges: when evidence is systematically collected, interpreted, and connected to clear decisions, it becomes an engine for meaningful educational change. Programme teams can move from isolated module-level assessments to coherent, programme-wide conversations about what students are truly learning, under which conditions, and with what levels of equity and consistency.

Sustained improvement requires institutional commitment. This includes ensuring time for faculty calibration, supporting data management infrastructures, and recognising evaluation work as part of academic practice. When these enabling conditions are in place, outcomes evaluation strengthens the alignment between curriculum design, teaching practices, and evolving professional expectations.

Looking ahead, engineering programmes face growing demands related to digitalisation, sustainability, global collaboration, and ethical responsibility. These demands cannot be addressed through assessment alone; they require robust evaluation systems capable of integrating direct evidence of performance, stakeholder perspectives, longitudinal data, and contextual interpretation. Embedding such systems across curricula supports transparency, enhances student agency, and builds trust with accreditation bodies and society at large.

Ultimately, evaluation should be understood not as compliance, but as an opportunity: a cyclical process that encourages reflection, fosters innovation, and ensures that engineering education remains relevant, equitable, and future-ready.

Guidelines for transfer to other contexts

While the examples are situated in specific institutional settings, several transferable principles emerge. First, clarify the object of evaluation (skill, outcome, intervention, programme, or institution) before selecting tools. Second, combine direct and indirect measures to strengthen validity and avoid over-reliance on self-report data. Third, calibrate assessment criteria among faculty to enhance reliability and fairness. Fourth, embed feedback loops so that evaluation results inform curriculum decisions rather than remain descriptive reports. Finally, ensure alignment between evaluation data, accreditation requirements, and institutional quality assurance processes to promote sustainable implementation. These principles are adaptable across disciplinary, institutional, and national contexts, provided they are adjusted to local governance structures, resource constraints, and stakeholder expectations.

Resources and Further reading

In addition to the literature referenced in this chapter, we also suggest Borrego et al. (2009), Borrego and Henderson (2014), Falchikov and Goldfinch (2000), Kane (2013) Lang et al. (2022), Messick (1995), Patton (2017), and Pellegrino et al (2014) to deepen your knowledge on evaluating using assessment outcomes.

References

Alkin, M. C., & Christie, C. A. (Eds.). (2023). Evaluation roots: Theory influencing practice. Guilford Press.

ASEAN University Network. (n.d.). ASEAN University Network–Quality Assurance (AUN-QA). https://www.aunsec.org/discover-aun/thematic-networks/aun-q

Augusti, G. (2009). EUR-ACE: The European accreditation system of engineering education and its global context. In P. Gray & A. Patil (Eds.), Engineering education quality assurance: A global perspective (pp. 41–49). Springer Nature. https://doi.org/10.1007/978-1-4419-0555-0_3

Banta, T. W., & Palomba, C. A. (2015). Assessment essentials: Planning, implementing, and improving assessment in higher education. John Wiley & Sons.

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). McGraw-Hill Education.

Borrego, M., Douglas, E. P., & Amelink, C. T. (2009). Quantitative, qualitative, and mixed research methods in engineering education. Journal of Engineering Education, 98(1), 53–66. https://doi.org/10.1002/j.2168-9830.2009.tb01005.x

Borrego, M., & Henderson, C. (2014). Increasing the use of evidence-based teaching in STEM higher education: A comparison of eight change strategies. Journal of Engineering Education, 103(2), 220–252. https://doi.org/10.1002/jee.20040

Chans, G. M., Valle-Arce, A. P., Salas-Maxemín, S., Caratozzolo, P., & Camacho-Zuñiga, C. (2025). Exploring transversal competencies in engineering students through international experiences. Frontiers in Education, 9, Article 1457796. https://doi.org/10.3389/feduc.2024.1457796

Christie, C. A., & Alkin, M. C. (2023). An evaluation theory tree: A framework for guiding a range of evaluation practice. In M. C. Malkin & C. A. Christie (Eds.), Evaluation roots: Theory influencing practice (pp. 13–65). Guilford Press.

Crawley, E. F., Malmqvist, J., Östlund, S., Brodeur, D. R., & Edström, K. (2014). Rethinking engineering education: The CDIO approach (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-05561-9

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of Educational Research, 70(3), 287–322. https://doi.org/10.2307/1170785

Felder, R. M., & Brent, R. (2003). Designing and teaching courses to satisfy the ABET engineering criteria. Journal of Engineering Education, 92(1), 7–25. https://doi.org/10.1002/j.2168-9830.2003.tb00734.x

International Engineering Alliance. (2021). Graduate attributes and professional competencies (4th ed.). https://www.internationalengineeringalliance.org/about/documents

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1–73. https://doi.org/10.1111/jedm.12000

Lang, C., Siemens, G., Wise, A. F., & Gašević, D. (Eds.). (2022). The Handbook of learning analytics (2nd ed.). SoLAR/Routledge.

Lönngren, J., & Svanström, M. (2016). Systems thinking for dealing with wicked sustainability problems: Beyond functionalist approaches. In W. Leal Filho & S. Nesbit (Eds.), New developments in engineering education for sustainable development (pp. 151–160). Springer. https://doi.org/10.1007/978-3-319-32933-8_14

Lucena, J., Downey, G., Jesiek, B., & Elber, S. (2008). Competencies beyond countries: The re-organization of engineering education in the United States, Europe, and Latin America. Journal of Engineering Education, 97(4), 433–447. https://doi.org/10.1002/j.2168-9830.2008.tb00991.x

Mahalingam, T. (2026). Program learning outcome assessment: a scalable framework for quality assurance in outcome-based education. Quality Assurance in Education, 34(2), 360–377. https://doi.org/10.1108/QAE-07-2025-0215

Messick, S. (1995). Validity of psychological assessment. American Psychologist, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. Academic Medicine, 65(9), S63–S67. https://doi.org/10.1097/00001888-199009000-00045

Muzulon, N. Z., Resende, L. M., Leal, G. C. L., & Pontes, J. (2025). Beyond technical skills: Competency framework for engineers in the digital transformation era. Societies, 15(8), Article 217. https://doi.org/10.3390/soc15080217

Olds, B. M., Moskal, B. M., & Miller, R. L. (2005). Assessment in engineering education: Evolution, approaches and future collaborations. Journal of Engineering Education, 94(1), 13–25. https://doi.org/10.1002/j.2168-9830.2005.tb00826.x

Osipovskaya, E., Coelho, A., & Tasi, P. (2024). Transversal competencies assessment and pedagogical methods for higher education: A literature review. In J. C. Rodrigues, R. Zimmermann, G. Dalmarco, & A. Simoes (Eds.), Human-centred technology management for a sustainable future, vol. 1 (pp. 189–197). Springer. https://doi.org/10.1007/978-3-031-72486-2_20

Patton, M. Q. (2017). Principles-focused evaluation: The guide. Guilford Publications.

Pellegrino, J. W., DiBello, L. V., & Brophy, S. P. (2015). The science and design of assessment in engineering education. In A. Johri & B. M. Olds (Eds.), Cambridge handbook of engineering education research (pp. 571–598). Cambridge University Press. https://doi.org/10.1017/CBO9781139013451

Sanchez-Carracedo, F., Sabate, F., & Gibert, K. (2021a). A methodology to assess the sustainability competencies in engineering undergraduate programs. International Journal of Engineering Education, 37(5), 1231-1243. https://doi.org/10.1016/j.jclepro.2021.126322

Sánchez-Carracedo, F., Ruiz-Morales, J., Valderrama-Hernández, R., Muñoz-Rodríguez, J. M., & Gomera, A. (2021b). Analysis of the presence of sustainability in higher education degrees of the Spanish university system. Studies in Higher Education, 46(2), 300–317. https://doi.org/10.1080/03075079.2019.1630811

Sánchez-Carracedo, F., Sureda, B., Moreno-Pino, F. M., & Romero-Portillo, D. (2021c). Education for sustainable development in Spanish engineering degrees. Case study. Journal of Cleaner Production, 294 (17), Article 126322. https://doi.org/10.1016/j.jclepro.2021.126322

Sánchez-Carracedo, F., Segalas, J., Busquets, P., Camacho, S., Climent, J., Lazzarini, B., Martín, C., Miñano, R., De Cámara, E. S., Sureda, B., Tejedor, G., & Vidal, E. (2022). Using competency maps for embedding and assessing sustainability in engineering degrees. Trends in Higher Education, 1(1), 58–81. https://doi.org/10.3390/higheredu1010006

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. Journal of the Learning Sciences, 23(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Sangwan, D., Sangwan, K. S., & Raj, P. (2022). 21st-century competencies in engineering education: initiation, evolution, current, and now whither to. In Proceedings 50th SEFI Annual Conference. (pp. 672–681). https://doi.org/10.5821/conference-9788412322262.1409

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining. In LAK '12: Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252–254). https://doi.org/10.1145/2330601.2330661

Stufflebeam, D. L., & Shinkfield, A. J. (2007). Evaluation theory, models, and applications. Jossey-Bass.

AI statement

During the preparation of this work, the authors used ChatGPT, Claude (Anthropic), Microsoft Editor and Grammarly to support the drafting, restructuring, shortening, and refinement of some sections of the chapter, particularly to improve clarity, flow, and practitioner-oriented wording. After using these tools, the authors reviewed, edited, and made the content their own as needed. They take full responsibility for the content of the publication.

Contributor statement

Conceptualisation	Adesola Adetunji, Ester Gimenez-Carbo, Irene Josa, Franch Maverick A. Lorilla
Data curation	Irene Josa
Investigation	Ester Gimenez-Carbo
Methodology	Adesola Adetunji, Ester Gimenez-Carbo, Irene Josa, Franch Maverick A. Lorilla
Project administration	Ester Gimenez-Carbo
Visualisation	Irene Josa, Adesola Adetunji, Ester Gimenez-Carbo
Writing - original draft	Adesola Adetunji, Ester Gimenez-Carbo, Irene Josa, Franch Maverick A. Lorilla
Writing - reviewing and editing	Adesola Adetunji, Ester Gimenez-Carbo, Irene Josa, Franch Maverick A. Lorilla