5.4

Evaluating Interventions That Develop Transferable Competencies and Skills

Joelyn de Lima[a], Siara Isaac[b], Oxana Sachenkova[c], Rucha Joshi[d], and Erna Engelbrecht[e]

[a] EPFL, Switzerland, ORCID: 0000-0001-9235-9704;
[b] EPFL, Switzerland, ORCID: 0000-0002-1527-8510;
[c] Linnaeus University, Sweden, ORCID: 0000-0002-3238-6011;
[d] Plaksha University, India, ORCID: 0000-0002-0456-233X;
[e] Delft University of Technology, the Netherlands, ORCID: 0000-0001-8731-347X.

Keywords: *course assessment, educational interventions, programme evaluation, reflective practice, designing evaluations, evaluating evaluations*

This chapter presents practical guidelines for educators to evaluate educational interventions designed to develop engineering students' transferable competencies and skills. The guidelines can be used when implementing new initiatives at the module or programme level, as well as documenting impact for your promotion dossier. Evaluation should be seen as an integral part of improving teaching and learning because documenting effective educational practices enables them to be appreciated and shared, and to identify points for improvement that fall short of your desired teaching goals. We provide a semi-sequential eight-step guide for articulating evaluation goals, identifying the intervention and skills to be evaluated, selecting methodologies and evaluation tools, and finally acting on the results of the evaluation. Additionally, this chapter offers criteria for assessing the quality of the evaluation itself to ensure the validity, reliability, and credibility of findings. Overall, the chapter will introduce you to tools to design evaluations to inform pedagogical decision making and strengthen the impact of interventions that develop transferable competencies and skills.

DOI: 10.59490/mt.266.97.

1 Introducing interventions for developing transferable competencies and skills

Educational interventions aimed at developing students' transferable competencies and skills, such as communication, teamwork, or critical thinking, are a key outcome for any well-designed engineering programme. This makes it vital to assess how effectively our teaching and educational interventions contribute to this critical aspect of engineering graduates' competence. Yet, many educators face a familiar question: how do we know whether our interventions are actually working? Evaluation is not just a bureaucratic requirement, but a cornerstone for understanding the impact of interventions, improving educational practices, identifying areas for refinement, and guiding future action (Rossi et al., 2019). It helps us move beyond relying on personal experience and intuition, toward methods proven effective through research. This process is inherently iterative and relies on feedback loops to improve or scale promising practices, ensuring that interventions genuinely achieve their intended outcomes.

At the level of individual modules, evaluation helps educators refine learning activities and assessment practices. At the programme level, it supports educators to build coherence across modules and alignment with accreditation requirements. For educators with responsibilities at the level of a programme, evaluation findings can inform strategic priorities and resource allocation. Across these levels, evaluation can address different kinds of outcomes, from student responses to longer-term behavioural and organisational impact, as reflected in models such as Kirkpatrick's (2016) evaluation framework, which we will explain in the next section.

We define an educational intervention as any modification to a programme, module, or instructional setting that has an impact on learning outcomes. Examples are the addition of a short workshop on time management to a project course, revision of the grading criteria for an internship, or inviting an expert from industry to co-teach. Depending on its scale and scope, some evaluations may be described as programme evaluations when they span multiple modules or student cohorts. Whatever the scale, educators need timely feedback to refine intervention strategies and improve learning. An evaluation seeks to "*describe and explain experiences of students and educators and to make judgements and [interpret] their effectiveness*" (Edwards, 1991, p. 13); that is, an evaluation involves both collecting relevant observations and making pedagogical decisions based on their analysis.

Understanding the nature of transferable competencies and skills is also key to their effective evaluation. Unlike purely conceptual knowledge, transferable skills encompass the contextual implementation of emotional, cognitive, and often practical domains. Their development is supported by educational experiences that provide opportunities for deliberate practice, explicit focus on transfer to new contexts, and consistent, actionable feedback. Educational interventions for developing engineering students' transferable skills vary in their scope, duration, and underlying design. Considering the different scales of educational interventions is useful for both designing effective interventions and selecting appropriate evaluation methodologies.

We can broadly categorise interventions by their temporal dimension and systemic nature as shown in Table 5.4.1.

Table 5.4.1: Examples of interventions for developing students' transferable competencies and skills, organised by temporal scale and systemic nature

Temporal scale (duration) → *Systemic nature ↓*	*Short-term, focused activities (hours to days)*	*Medium-term engagements (week to a semester)*	*Long term*
Individual and module-level changes; affect individual students or module deliveries	Targeted workshops, modules, or active-earning activities like those from 3T Play (Isaac & de Lima, 2024).	Structured mentoring sessions, dedicated project work within modules, or formal internships.	Changes in instructor profiles (e.g., industry professionals as co-educators).
Programmatic and institutional changes; broader interventions reshaping the educational experience	Funding large-scale interdisciplinary student-run projects (e.g., iGEM, hyperloop, or solar car or boat competitions).	Curriculum modifications, e.g., structure of internships across curricula, introduction of new cross-departmental degree programmes, integrating makerspaces or other external collaboration into modules	Curriculum redesign: real-client projects, shifting internships from industry to civil society and social agencies

Given the diversity of educational interventions that develop transferable competencies and skills, it can be challenging to determine appropriate ways to conduct relevant evaluations. For other approaches to evaluation and research of transferable skills and competences we refer you to Chapter 5.1 and for more on meaningful educational change we refer to Fullan (2007).

This chapter is intended to give you, the educator, practical guidance on how to both design and interpret an evaluation of interventions targeting students' transferable competencies and skills. We briefly review the design of educational interventions that can facilitate their development. We argue that the evaluation of such interventions should inform their very design, allowing for greater coherence between conceptual understandings of skill development and the implementation of educational interventions. We will continue with a balance of practical considerations and

conceptual framing to prompt you to reflect on the motivations for conducting evaluations, to design and conduct relevant evaluations, and to interpret and benefit from the results of evaluations to improve your interventions for teaching transferable competencies and skills to engineering students

2 The context of evaluating interventions that develop transferable competencies and skills

When evaluating interventions for developing engineering students' transferable skills, it is important to consider why you want to do the evaluation and what its possible consequences may be. Besides the educational aims the intervention, it is important to note that other considerations may shape the evaluation, such as your own personal motivation (e.g., preparing a tenure application or compliance with funding requirements), the needs of any other stakeholders (e.g., department chair, future students, funding agency), and the potential implications of the findings. Given this, it is helpful to start by clearly defining the core question you want your evaluation to answer.

Evaluation also plays a key role in aligning your educational interventions with institutional goals and strategic objectives (Pohlenz, 2022), such as student success and accreditation compliance. Evaluation findings can influence resource allocation, funding, and sustained support. Rigorous evaluation can significantly boost eligibility for grants and funding by demonstrating tangible success, such as increased student retention.

Creating evaluations inherently involves subjectivity and bias, as they are shaped by the evaluator's role, context, and perspectives. Recognising these inherent judgements is essential, as they can influence how findings are interpreted and utilised. One significant example of this is reporting bias, where authors may be inclined to self-censor results or decide not to publish certain works (Pigott et al., 2013). This bias often leads to a disproportionate publication of positive results that indicate the success of an intervention, creating a skewed distribution of research knowledge and a potentially misleading impression from the literature (Dawson & Dawson, 2018).

The evaluation process can provide insight into how and why an intervention works, or conversely, why it might fail. Effective interventions for developing transferable competencies and skills are built on a clear understanding of their intended learning outcomes and the pedagogical conditions appropriate to realise these outcomes. This means an evaluation should start with a clear statement of the learning objectives you are aiming for, how your activities are designed to support those outcomes, and how you will assess whether students have achieved them.

Considering these key ideas about the purpose of evaluation, the nature of interventions, and the need to understand how an intervention impacts learners, we turn our attention to the practicalities of how to evaluate interventions for transferable skills. We will introduce conceptual frameworks and present some guidelines, illustrated through concrete examples to highlight common pitfalls.

Evaluation frameworks such as Kirkpatrick and Kirkpatrick's (2016) can help educators clarify whether they are primarily interested in students' experiences (reaction — level 1), learning (level 2), behaviour change (level 3), or broader impact (results — level 4) of the module. These four levels offer a systematic approach to understanding effectiveness of programmes and modules. The two most accessible levels of this model are *reaction*, which seeks to capture participants' impressions and experiences, and *learning*, which involves assessing their acquired knowledge and skills. Taking the example of transferable competencies and skills for a constructive team climate that includes psychological safety, the *reaction* level could capture if students appreciated an educational intervention (e.g., I liked, I saw value in, I preferred, etc.). The *learning* level measures if the intervention increased what students knew (i.e., assessing knowledge gain) about constructive team climate. The level of *behaviour* could document what participants did to support psychological safety in their project teams. The *results* level documents the impact of these new behaviours on participants' other activities and accomplishments, such as the quality of students' project work or sense of team belonging improved.

Designing evaluations at Kirkpatrick levels 3 and 4 is considerably more complex due to the extended time frames and higher resource requirements. However, these upper levels typically align more closely with the desired impacts for teaching interventions. It is therefore encouraging to note that capturing students' intention to use course material serves as a robust indicator of their future use of course material (Alliger et al., 1997). Indeed, Alliger and colleagues found that students stating that course material was useful to them was a better predictor of them using it (level 3) than direct measures of their learning (level 2).

Complementing this, Brookfield's (1995) framework for reflective practice helps educators interpret evaluation evidence by deliberately incorporating multiple perspectives (lenses): the educator's own experiences, student feedback, colleagues' insights, and the theoretical lens of scholarly literature. A robust evaluation approach can assist educators to engage productively with all lenses, particularly that of scholarly literature which is less familiar to them.

For example, module evaluation questionnaires frequently used in higher education give us insight into student perspectives (both a Brookfield lens and an example of Kirkpatrick and Kirkpatrick's level 1). While valuable, the utility of student feedback can be obscured by poorly constructed questionnaires that are, for example, designed for lecture modules rather than project-based modules, with questions that focus too much on educator activities. Another, perhaps larger issue, is that the student requests are often incorrectly interpreted as sound pedagogical recommendations. Brookfield's approach requires that student opinion questionnaires be digested and analysed from multiple perspectives, including literature. For example, students may state that they would prefer a short lecture on the theory of psychology safety without an application activity. Analysing this student feedback from a learning sciences perspective reminds us that skill development is improved by low-stakes application activities with immediate feedback, so a more constructive response could be preparing students for

the application activity by informing them about the mechanisms of skill development. Consider discussing student feedback with an education expert, as this has been shown to improve teaching (Marsch & Roche, 1993). The challenge of more integrated approaches to evaluation is that they take time, expertise, and coordination. We propose some practical guidelines to accompany you through the conceptualisation, development, and implementation of a coherent evaluation.

3 Approaches to evaluating interventions in skills education

3.1 A practical overview of how to evaluate interventions that develop transferable competencies and skills

This section provides practical recommendations for evaluating interventions that develop transferable competencies and skills. As the approach and methods should be informed by your goals and context, our recommendations start with clarifying the goal of your evaluation. Fig. 5.4.1 provides a semi-sequential series of steps for preparing an evaluation of an intervention that teaches transferable skills. As ever, some recursive revision is likely as your project advances. When the evaluation is designed ahead of time, you can integrate the evaluation into the intervention itself, taking advantage of moments when you are already interacting with your target group.

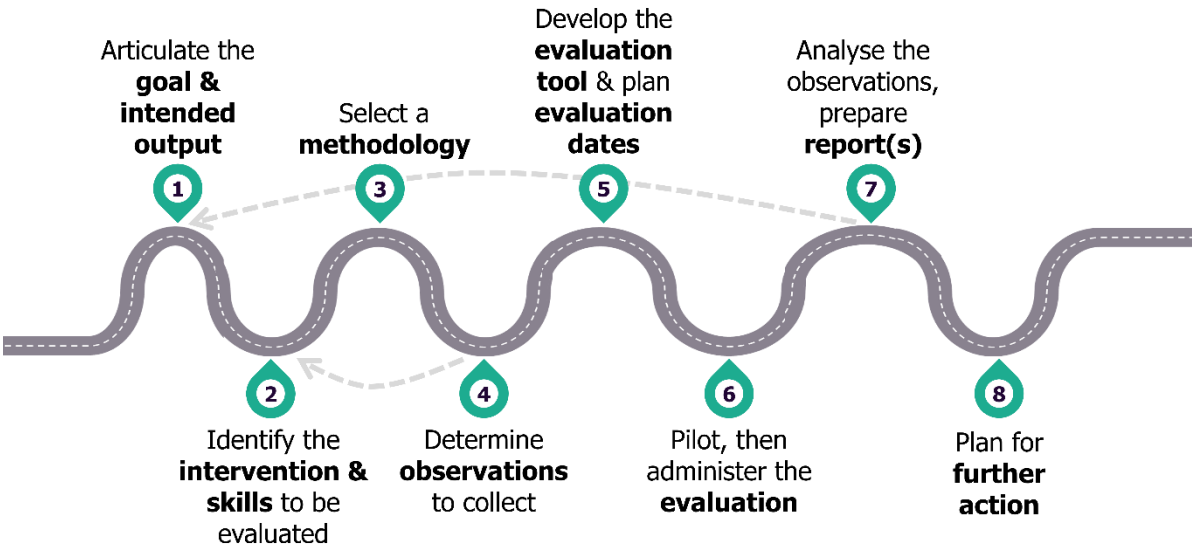


Figure 5.4.1: A semi-sequential checklist for preparing an evaluation. Own Work by de Lima

Step 1. Start by clarifying why you are conducting an evaluation, including how the results will be documented to reach your intended audience. Depending on your motivations, relevant outputs might include an internal report, a conference paper, or your tenure committee. Having a concise record of your goals assists in maintaining a coherent approach.

It is important to distinguish between assessment of learning and evaluation of an intervention. In general, assessments are conducted to determine students' performance in a module, while the objective of evaluation is to improve the intervention. Evaluation focuses on understanding how and why an educational intervention contributed to student learning and other outcomes. Many methods used in assessment can also be used in evaluation, and one can support the other.

Step 2. Next, clearly specify the educational intervention being evaluated. Due to expectations that these skills are implicitly developed in higher education (Dokter, 2024; Isaac et al., 2023), and since they are often described too broadly (e.g., communication skills), it may be necessary to describe the transferable skills with adequate specificity to allow for meaningful evaluation. You could use the Activity Design Worksheet A by Isaac and de Lima (2024, pp. 195-197) to guide your process of clarifying the specific skills and granularity that will be your focus. When describing the intervention to be assessed, specify the time frame, participants, and other elements that are relevant to planning.

Step 3. There are multiple methodological options for evaluating interventions that develop transferable skills. The appropriateness of each depends on your evaluation goal, for example whether you aim to understand student experiences, document behavioural change, or compare alternative educational designs. Before deciding, we recommend considering a couple of options:

- *Participatory strategies,* such as engaging with students as partners, or collecting observations from project supervisors or future employers, can bring in rich perspectives.
- *Qualitative methods,* such as interviews, can help you explain how and why an intervention works.
- *Quantitative methods,* which are useful when you want to compare groups or track changes over time.

Selecting a method involves determining both the measure that will enable you to assess the impact of the intervention and the implementation approach. Large-scale questionnaires may seem most accessible but should be used with consideration after exploring alternative methodologies in the literature. Two final elements to consider in selecting your method are how you will reach your desired population for data collection (obtaining sufficient representativity, diversity, or quantity depending on your methodology) and what resources are needed for collecting, cleaning, and analysing data (distribution, transcription, etc.). A decision grid that lists possible approaches along with their associated benefits and constraints can help you to make well-informed decisions. Revisit and refine your response to step 1 to ensure your approach stays aligned with your goals.

Step 4. Decide what data you will collect, from which stakeholders, and on which aspects. For example, to compare students' perceptions of transferable skills development in client-based projects versus more traditional academic ones, it may be more effective to survey students with experience in several projects. Other

potential approaches include documenting educators' experiences with implementation, comparing costs, analysing transferable skill learning outcomes across the curriculum, observing how teaching assistants interact with students, or tracking participation across student profiles. If you conclude that assessing students' skills themselves is most appropriate, we recommend turning to Part 3 of this handbook.

Step 5. You are ready to start developing the evaluation instrument, i.e., the questions, task, or situation that will allow you to gather the data you need. Using a previously developed instrument, such as a validated questionnaire or set of interview questions, strengthens your study by providing conceptual coherency and a potential baseline. These are important benefits that are unfortunately underutilised in education (Buntins et al., 2021). In addition to developing an instrument to elicit relevant data, you must consider *when* to collect your data to best capture relevant information about transferable skills.

In parallel, plan logistics such as dates, physical or digital resources, and clear information to encourage your target audience to participate. Verify if your institution and/or intended publication venue require ethical approval and account for potential delays in obtaining approval. In all cases, be mindful of ethical considerations and only collect the necessary data to accomplish your evaluation goals and ensure participants have consented to participate (informed consent).

Some examples of data collection instruments are:

- *Self-assessment questionnaires* targeting the learning experience (e.g., based on the 3T PLAY framework for experiential interventions; Isaac & de Lima, 2024, p. 17 and Chapter 2.10) or targeting different skill categories (e.g., at the end of a project-based module; Tormey & Laperrouza, 2023) and other methods as discussed in Chapter 3.2 on self-assessment.
- *Classroom observations* either using video recordings or by asking a colleague to observe your teaching (Guimarães & Lima, 2021).
- *Interviews and focus groups* with your students.
- Analysing *student artifacts* such as reports, reflections, online discussion posts, learning journals, or portfolios.
- *Digital learning analytics* or trace data that measures online engagement and submission patterns in your online course environment.

Step 6. Conduct a pilot evaluation to identify potential problems. Piloting saves time by ironing out issues before data collection and can often be done while waiting for ethical approval. Test both technical and conceptual issues under conditions as close as possible to those planned for your implementation. Our personal experiences of off-campus students being unable to open hyperlinks, too much background noise, and video file formats being incompatible with data processing software illustrate the benefits of piloting every technical step. Conceptual piloting ensures a shared understanding of the instrument between yourself, your participants, and your colleagues. This can involve focus group discussions (to test your quantitative

instruments) or a test interview with a colleague, followed by transcription and analysis (including inter-rater reliability analysis with colleagues) using your intended tools. Build in flexibility by planning extra time, data collection dates, additional participants, etc. to ensure that last-minute technical issues or incomplete sessions do not derail your entire evaluation.

Step 7. When starting your analysis, ensure that your analysis is focused and grounded by reminding yourself of the original questions and intended outcomes. Applying a relevant conceptual framework throughout the process will help you interpret your findings with greater depth and provide you with a clearer perspective on the implications of your results. Consider re-engaging with your stakeholders, such as your research participants or your intended audience, during your analysis phase to enrich your interpretation. When writing your results, keep your initial objective and your target audience in mind. It may be most effective to prepare two documents, such as a journal article and an internal report, each tailored to the needs of that audience. Avoid bias, such as underreporting negative findings or overstating positive outcomes (Cohen et al., 2018; Miles et al., 2020). We refer you to Chapter 5.6 for further information on dissemination of your results.

Step 8. The final step is to consider what actions should be undertaken based on your findings. This can include ideas for improving this specific intervention, additional interventions, or conducting another evaluation. Clear recommendations increase the value of your evaluation and increase the likelihood of constructive implementation. When communicating your results, be strategic and align your findings with institutional goals, when possible (Soliman et al., 2019), while being mindful of reporting bias (Dawson & Dawson, 2018). This can strengthen your case, for instance when advocating for funding. When preparing communications for policy makers or academic audiences, take a scholarly lens by contextualising with literature, including critical perspectives and implications. Your reporting should allow your audiences to make decisions, so be explicit in your recommendations. For guidance on how to use your evaluation to maximum effect, see Section 5.

3.2 Integrative approaches for evaluating interventions

Over the years, many methods of evaluating teaching interventions have emerged, such as student ratings, peer or expert reviews, and video analysis, each offering a different perspective. What is important is not just measuring learning outcomes but also capturing the context, processes, and lived experiences that shape them.

Self-reporting by teaching staff is one way of doing this. Such reflective narratives often highlight why an intervention was tried, the strategies used, and how the module evolved based on student feedback. Compiled into teaching portfolios (Seldin, 1993), they provide rich insights and remain a valuable tool for both documenting practice and driving continuous improvement (Braskamp & Ory, 1994).

When we look at teaching interventions, it is not enough to merely ask 'does it work?' We also need to ask how and why it works, and under what conditions. For a while, the evidence-based movement (Hammersley, 2007; Hanley et al., 2016; Pring &

Thomas, 2004) leaned heavily on randomised controlled trials (RCTs), because they gave clear, generalisable results (Connolly et al., 2018; Torgerson & Torgerson, 2003). However, critics have pointed out that while RCTs are useful to establish replicable outcomes, this reductionist approach fails to reflect the complexity of real-world educational settings (Connolly et al., 2018; Outhwaite et al., 2020). By focussing on statistical generalisation, such methods can disconnect interventions from the contexts in which they are embedded (Elliott, 2001). Instead, we need to pay closer attention to how interventions are implemented and evaluate the process of implementation through a holistic approach.

Researchers also caution against a heavy reliance on self-report questionnaires, varied study designs (many lacking control groups), and an overall absence of evidence demonstrating real-world transfer (Tyne et al., 2024). These issues reflect wider challenges in evaluation, such as insufficient mapping of interventions to learning objectives, over-reliance on surveys, lack of baseline data, unbalanced use of quantitative versus qualitative data, poorly defined evaluation criteria, and missing justifications grounded in evidence or theory. A more integrated approach is required to solve such issues.

Another tricky challenge is figuring out how to evaluate multiple goals of an intervention at once. Take the example of project-based learning: you might want to know if it boosted student performance, improved client satisfaction, and motivated students and educators — all at the same time. To handle this complexity, we need structured ways to capture the bigger picture. Venezky (2001) suggested two useful tools. One is the History of the Future exercise, where you imagine the intervention fully succeeding and draft a future press release about what was achieved, how it worked, and why it mattered. The other is Programme Logic Mapping, which visually connects inputs (like the intervention) to outputs (such as better engagement) and intermediate outcomes (like educator behaviour changes). Together, they help clarify goals, test assumptions, and create a more coherent evaluation.

4 Implications for evaluating interventions that develop transferable competencies and skills

It is important to consider the quality of the evaluation itself. Evaluations demand resources from multiple stakeholders (e.g., students, instructors, institutions) and can have significant consequences for the future of the intervention and institutional goals and credibility. Poorly designed evaluations risk pedagogical misalignment, misallocated resources, damage to the institution's reputation, and erosion of trust between the various stakeholders including accreditation bodies. However, well-designed and implemented evaluations can improve learning locally and support larger-scale adoption.

This section is designed to help you as an educator to evaluate your interventions, or to assess the quality of an evaluation done by another educator. It briefly introduces criteria that can be used for this purpose, provides practical ideas and a checklist (Fig. 5.4.2), and uses two mini case studies of educators that operationalise the evaluation

quality criteria in interventions that develop transferable competencies and skills to help you contextualise the criteria (Tables 5.4.2 and 5.4.3). There are several resources that can provide you with further support and details for the quality of evaluation (e.g., American Educational Research Association [AERA] et al., 2014; Cruz et al., 2019; Gullickson, 2003; Messick, 1995; Organisation for Economic Co-operation and Development, 2021; Yarbrough et al., 2010).

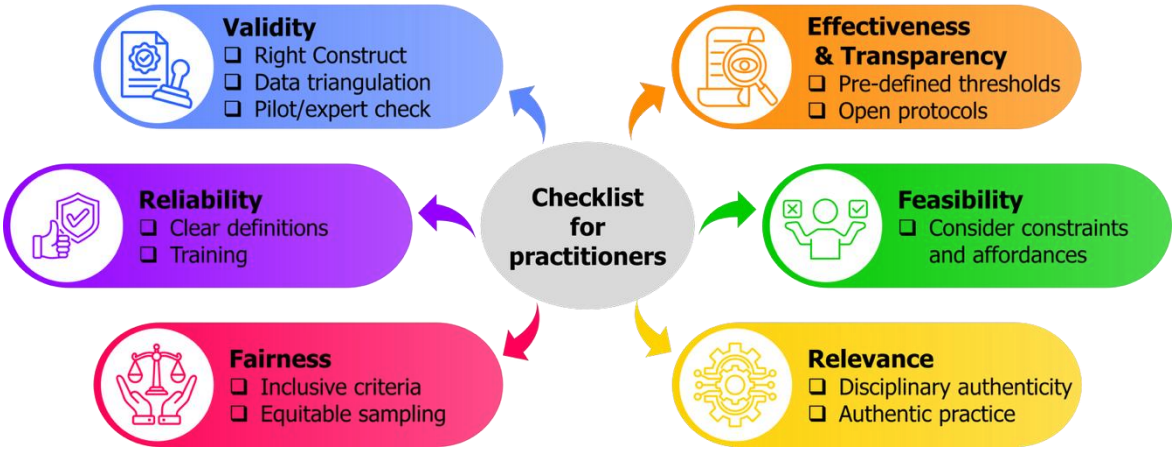


Figure 5.4.2: Checklist for practitioners who seek to evaluate the quality of evaluations. Own Work by de Lima

Validity considers if the data genuinely say what you think it does. While validity can include several aspects, we focus on this key question: do you trust the claims or inferences made from the evaluation results? To assess validity, ask:

- Are you measuring what you intended to measure? Staying aligned with your goals will prevent getting sidetracked by irrelevant details.
- Are you using multiple sources of data? Considering multiple stakeholders, perspectives, and types of data will allow for triangulation and strengthen the credibility of your findings.
- Did you pilot your instruments, or have an expert review them? This will help early detection of potential issues with validity, gaps in the data, or unintentional outcomes.

Reliability considers the consistency of the data obtained in the evaluation process and focuses on how data is collected and analysed. This means that if other people conducted the evaluation, they would reach similar results. To assess reliability, ask:

- Do you have clear operational definitions or well-defined rubrics? Clear criteria reduce ambiguity and ensure that everyone measures the same thing in the same way.
- Are data collectors and analysts trained and following the same procedures? This ensures consistency and reduces variation caused by individual judgements or different approaches.

A fair evaluation explicitly accounts for diversity in individuals and contexts and actively mitigates potential biases. To assess fairness, ask:

- Do the benchmarks account for diverse participants and ways of participating? This avoids systematically marginalising or privileging certain groups.
- Are your sampling and analysis equitable? Collecting data across different times and groups while minimising unnecessary burden on the students/observers/analysers supports fairness.

Feasibility involves being realistic about the resources available in your context, especially when replicating an evaluation done in a different context. While integrating evaluation into the intervention design can assist, assess feasibility by asking:

- Have you considered each part of the evaluation alongside your affordances and constraints? Aligning design, data collection, and analysis with available expertise, time, and tools ensures feasibility.
- If the original plan is not feasible, what is the most you can do with available resources? You may need to adjust your evaluation goals.

The evaluation should be *relevant* for your teaching or your institutional goals and preferably align with broader disciplinary and global needs. In engineering education, relevance means the questions, methods, and claims are authentic to the discipline and linked with professional standards, ideally informing accreditations, curriculum design, or graduate career trajectories. To assess relevance, ask:

- Is the goal aligned with industry needs or accreditation standards?
- Does the evaluation mirror authentic disciplinary practices?

Effectiveness determines if an evaluation meets its intended goals. Therefore, setting clear goals and targets at the outset is essential. However, results should be interpreted in context, considering the possible factors that contribute to differential results for sub-populations, and the relative importance they hold for different stakeholders. To assess effectiveness, ask:

- Did you set clear goals and targets, and did the evaluation provide credible evidence to demonstrate achievement? Strong evidence should withstand peer scrutiny.
- Are the results different for any sub-populations, and can you determine which factors most influenced the outcomes? Understanding these differences will support meaningful interpretation.

Transparency means giving visibility to typically invisible aspects, so that stakeholders can understand, trust, and if necessary, critique the evaluation. It is important to get a nuanced understanding of an evaluation. This is well established in fields like medicine (Cotterill et al., 2018) and increasingly called for in higher education (Shavelson, 2010). To assess transparency, ask:

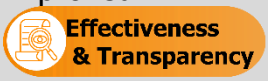
- Is the process of gathering and analysing data well documented? Clear records allow peers to understand exactly what was done, and why.

- Have you been explicit about the limitations of your method, data, analysis, or claims? Acknowledging constraints adds credibility, not weakness.
- Does the reader understand your role and motivations? Knowing your position helps the reader interpret your perspectives.

To help you contextualise and adopt the criteria we have used two mini-cases of educators that operationalise evaluation quality criteria in interventions that develop transferable competencies and skills (Table 5.4.2).

Table 5.4.2: Mini case studies illustrating strategies that improve validity, reliability, fairness, feasibility, relevance, effectiveness, and transparency in the evaluation of teaching transferable competencies and skills.

Context	*Case 1: an educator implemented a short active learning activity in their module designed to improve communication within teams.*	*Case 2: an educator was asked to review their department's internships for an upcoming accreditation.*
Goal of evaluation	They want to *measure students' behavioural engagement* during the activity.	They need to *assess the impact of internships on students' transferable skill proficiency*.
Strategies that improved **Validity**	The educator collected data about actual behavioural engagement using a validated protocol.	The educator collected the students' reports and implemented a pre- and post-internship questionnaire data assessing students' skill proficiency, from multiple sources (students, supervisors, and a control group of non-participants).
Strategies that improved **Reliability**	The educator used a rubric with clear definitions, e.g., only count when a student verbally clarifies a concept to a peer.	The educator trained raters by having them grade five anonymised reports to compare scores and resolve discrepancies.
Strategies that improved **Fairness**	The educator asked two colleagues to be observers; they focussed on multiple behaviours, not just verbal interactions, to avoid disadvantaging students with fewer verbal skills.	The educator gathered multiple types of data from multiple stakeholders.

Context	*Case 1: an educator implemented a short active-learning activity in their module designed to improve communication within teams.*	*Case 2: an educator was asked to review their department's internships for an upcoming accreditation.*
Goal of evaluation	They want to measure students' behavioural engagement during the activity.	They need to assess the impact of work-based internships on students' transferable skill proficiency.
Strategies that improved 	The educator determined that their team does not have expertise to develop a new instrument but has expertise in quantitative data analysis skills. So, they could use a previously validated protocol with a strong quantitative component.	The educator determined that the report templates and questionnaires already existed, and they could expect a high response rate. However, they would need to budget for external qualitative data analysis.
Strategies that improved 	This evaluation is relevant as communication in teams is both an authentic disciplinary practice and part of engineering programme accreditation criteria.	This evaluation is relevant as transferable skills are emphasised in accreditation criteria, and the findings could have implications for all institutions with internships.
Strategies that improved 	The educator's report showed a 25% increase in oral interactions, exceeding the 20% goal, with the caveat that online interactions were not captured. It included anonymised data and noted limitations (observer presence influencing student behaviour).	The educator's report demonstrated effectiveness, as supervisors' perceptions of student skill proficiency increased after the internship, and also included the questionnaire and limitations (potential conflict of interest as their departmental accreditation depends on demonstrating the effectiveness of the intervention).

5 Recommendations for teaching practice

Evaluating interventions that teach transferable competencies and skills should not rely exclusively on student opinion questionnaires or grades. Using multiple sources for data collection will help you make sense of how well your interventions meet your goals and why (or why not). Of course, the approaches discussed in this chapter should always be adapted to local contexts.

The true value of an evaluation lies in using the findings to improve how we teach transferable skills in engineering students. Bringing in different perspectives and triangulating your data can provide important and useful insights (Patton, 2015). When sharing the findings of your evaluation, consider both the intended audiences and the best way to support constructive engagement with the results. Different stakeholders value different types of information, so clarify how the findings should inform their decisions and what actions from them can lead to meaningful change. Colleagues, administrators, student representatives, or external stakeholders may need to examine and digest your findings before they will be ready to act on them. Rushing this phase can increase resistance to change, especially when findings oppose common practices or institutional norms. We refer to Chapter 5.6 for further information on dissemination of your results.

Ideally, evaluating interventions should not be treated as a once-off event but as part of the continuous improvement of our daily teaching practice. We encourage you to embed it into routine activities by drawing on evidence from student project work, reflective assignments, feedback moments, and classroom interactions. Systematic collection and review of this evidence aligned with existing teaching tasks can make the evaluation of interventions efficient and sustainable. For more information about the systematic use of assessment data to evaluate educational outcomes and improve programmes, see Chapter 5.3.

References

Alliger, G. M., Tannenbaum, S. I., Bennett Jr, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. Personnel Psychology, 50(2), 341–358. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). Standards for educational and psychological testing. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Braskamp, L. A., & Ory, J. C. (1994). Assessing faculty work: Enhancing individual and institutional performance. Jossey-Bass.

Brookfield, S. D. (1995). Becoming a critically reflective teacher. Jossey-Bass.

Buntins, K., Kerres, M., & Heinemann, A. (2021). A scoping review of research instruments for measuring student engagement: In need for convergence. International Journal of Educational Research, 2, Article 100099. https://doi.org/10.1016/j.ijedro.2021.100099

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. Educational Research, 60(3), 276–291. https://doi.org/10.1080/00131881.2018.1493353

Cotterill, S., Knowles, S., Martindale, A.-M., Elvey, R., Howard, S., Coupe, N., Wilson, P., & Spence, M. (2018). Getting messier with TIDieR: Embracing context and complexity in intervention reporting. BMC Medical Research Methodology, 18, Article 12. https://doi.org/10.1186/s12874-017-0461-y

Cruz, M. L., Saunders-Smits, G. N., & Groen, P. (2019). Evaluation of competency methods in engineering education: A systematic review. European Journal of Engineering Education, 45(5), 729–757. https://doi.org/10.1080/03043797.2019.1671810

Dawson, P., & Dawson, S. L. (2018). Sharing successes and hiding failures: 'Reporting bias' in learning and teaching research. Studies in Higher Education, 43(8), 1405–1416. https://doi.org/10.1080/03075079.2016.1258052

Dokter, L. (2024). Academics' perspectives on transversal skills: A case study in a new engineering programme at the Norwegian Defence University College. In Proceedings of the 52nd Annual Conference of SEFI, Lausanne, Switzerland. https://doi.org/10.5281/zenodo.14254766

Edwards, J. (1991). Evaluation in adult and further education: A practical handbook for teachers and organisers. Workers' Educational Association. https://unesdoc.unesco.org/ark:/48223/pf0000193933

Elliott, J. (2001). Making evidence-based practice educational. British Educational Research Journal, 27(5), 555–574. https://doi.org/10.1080/01411920120095735

Fullan, M. (2007). The new meaning of educational change (4th ed.). Routledge.

Guimarães, L. M., & Lima, R. D. S. (2021). A systematic literature review of classroom observation protocols and their adequacy for engineering education in active learning environments. European Journal of Engineering Education, 46(6), 908–930. https://doi.org/10.1080/03043797.2021.1937946

Gullickson, A. R. (2003). The student evaluation standards: How to improve evaluations of students. Corwin Press.

Hammersley, M. (2007). Educational research and evidence-based practice. SAGE.

Hanley, P., Chambers, B., & Haslam, J. (2016). Reassessing RCTs as the 'gold standard': Synergy not separatism in evaluation designs. International Journal of Research & Method in Education, 39(3), 287–298. https://doi.org/10.1080/1743727X.2016.1138457

Isaac, S., & de Lima, J. (2024). Teaching transversal skills for engineering. EPFL Press. https://doi.org/10.55430/SJYVJ3TP24

Isaac, S., Petringa, N., Jalali, Y., Tormey, R., & Dehler Zufferey, J. (2023). Are engineering teachers ready to leverage the power of play to teach transversal skills? In Proceedings of the 51st SEFI Annual Conference. https://doi.org/10.21427/QP3D-B914

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). Kirkpatrick's four levels of training evaluation. ATD Press.

Magana, S. (2017). Disruptive Classroom Technologies: A Framework for Innovation in Education. Corwin. https://doi.org/10.4135/9781071872628

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). Qualitative data analysis: A methods sourcebook (4th edition). SAGE.

Organisation for Economic Co-operation and Development. (2021). Applying evaluation criteria thoughtfully. OECD Publishing. https://doi.org/10.1787/543e84ed-en

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2020). A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. International Journal of Research & Method in Education, 43(3), 225–242. https://doi.org/10.1080/1743727X.2019.1657081

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). SAGE.

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. Educational Researcher, 42(8), 424–432. https://doi.org/10.3102/0013189x13507104

Pohlenz, P. (2022). Innovation, professionalisation and evaluation: Implications for quality management in higher education. Quality in Higher Education, 28(1), 50–64. https://doi.org/10.1080/13538322.2021.1951452

Pring, R., & Thomas, G. (2004). Evidence-based practice in education. McGraw-Hill Education.

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). Evaluation: A systematic approach (8th edition). Sage.

Seldin, P. (1993). Successful use of teaching portfolios. Anker.

Shavelson, R. J. (2010). Measuring college learning responsibly: Accountability in a new era. Stanford University Press.

Soliman, A. Y., Badawi, A. A., Elkordy, M. M., & Al-Bahi, A. M. (2019). A framework for constructing and assessing knowledge management systems for engineering institutes. In IEEE Global Engineering Education Conference (pp. 727–737). https://doi.org/10.1109/EDUCON.2019.8725176

Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. British Journal of Educational Studies, 51(1), 36–45. https://doi.org/10.1111/1467-8527.t01-2-00223

Tormey, R., & Laperrouza, M. (2023). The development, validation and use of an interprofessional project management questionnaire in engineering education. European Journal of Engineering Education, 48(3), 502–517. https://doi.org/10.1080/03043797.2023.2171854

Tyne, W. P., Fletcher, D., Paine, N. J., & Stevinson, C. (2024). Physical challenge interventions and the development of transferable skills for the workplace: A systematic review and meta-analysis. Organizational Psychology Review, 14(4), 618–667. https://doi.org/10.1177/20413866241245301

Venezky, R. L. (2001). Procedures for evaluating the impact of complex educational interventions. Journal of Science Education and Technology, 10(1), 17–30. https://doi.org/10.1023/A:1016664426024

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). The program evaluation standards: A guide for evaluators and evaluation users. SAGE.

AI statement

During the preparation of this work the authors did not make use of any AI tools.

Contributor statement

Supervision	Erna Engelbrecht
Visualisation	Joelyn de Lima
Writing - original draft	Joelyn de Lima, Siara Isaac, Oxana Sachenkova, Rucha Joshi, Erna Engelbrecht
Writing - reviewing and editing	Joelyn de Lima, Siara Isaac, Oxana Sachenkova, Rucha Joshi, Erna Engelbrecht