

CHAPTER 8

Quality and evaluation

This chapter examines the definitions of ‘quality’ that become operationalised as a suite of standards in both national and international contexts. These standards are a necessary consideration for providers of accredited qualifications and so the question of whether they are sufficient kitemarks for the more recent stackable microcredential qualifications is explored, together with their validation through the process known as ‘evaluation’.

How can we know quality when we see it?

Educators are usually confident about judging their students’ work and awarding the submitted assignments suitable marks.

How to cite this book chapter:

Ferguson, R. and Whitelock, D. 2024. Quality and evaluation. In: Ferguson, R. and Whitelock, D. *Microcredentials for Excellence: A Practical Guide*. Pp. 217–262. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bcz.h>. License: CC BY-NC 4.0

They can judge the value and the quality of students' performance of the task in hand, which is communicated as a mark. This process usually works well for formative assessment but with higher-stakes examinations, where double marking occurs, there is not always agreement. This is particularly true in the arts and even architecture, where ranking assignments has been found to lead to more agreement and provides a better metric for quality (van den Heuvel & Bohm 2023). Therefore, 'seeing' even for arts experts is not necessarily believing, which suggests one of the continuous problems around quality is not recognising it intuitively but employing an agreed, robust set of metrics or key performance indicators that can be used to review it systematically. Why is this important? The answer lies within a continuous improvement cycle of educational provision where the analysis of relevant metrics forms a foundation for quality advancement. Furthermore, with newer qualifications such as microcredentials, the elements of trust and transparency can be evidenced by quality assurance processes (Orr, Pupinis & Kirdulyte 2020).

There is also a political and ethical dimension to quality standards, as illustrated by Europe's aspiration to achieve the European Education Area by 2025 with 'high quality digital learning quality to increase the relevance quality of European education and Training' (European Commission 2020). The following section unpacks the notion of quality through a discussion of its measurement using both international and national standards.

International quality standards

A range of existing quality standards were initially designed for face-to-face teaching and learning and a number of quality assurance tools have been specifically developed to ensure the quality of online education, for example E-xcellence (see Rosewell et al.

2017). In addition, a range of stakeholders, not least students when choosing where to study, take note of ranking systems. However, Brasher et al. (2022) note that current ranking systems are of limited value for most potential undergraduate students, particularly with reference to online education, as these systems have been slow to include online teaching metrics into their analytics. These are valid points to bear in mind when microcredentials are delivered online. It is important to examine the basic quality frameworks and standards that already exist before discussing the quality recommendations of microcredentials, as these are being produced by higher education establishments that comply with existing regulatory guidelines.

A clear generic example is provided by the European Association for Quality Assurance in Higher Education (ENQA); a set of Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG 2015) is based upon four principles for quality assurance:

- higher education institutions have primary responsibility for the quality of their provision and its assurance;
- quality assurance responds to the diversity of higher education systems, institutions, programmes and students;
- quality assurance supports the development of a quality culture;
- quality assurance takes into account the needs and expectations of students, all other stakeholders and society.

These principles leave scope for individual circumstances and cultures, which can be reflected in the education policies of degree-awarding institutions. They also allow delegation of regulation and external quality assurance reviews to be undertaken by national bodies.

Moving on to unpacking these principles, 10 standards are proposed, as follows:

1. generation of a policy for quality assurance, which links to all the principles;
2. a design and approval process should be in place;
3. student-centred learning, teaching and assessment need to be explicit;
4. a process for student admission, progression, recognition and certification is required;
5. ensuring that teaching staff are competent;
6. learning resources and student support should be available;
7. information management should include the analysis of relevant data to maintain progress;
8. public information should be available;
9. programmes should be monitored on an ongoing basis and periodically reviewed;
10. there should be cyclical external quality assurance, usually undertaken by a national quality assurance agency.

All these standards are valid for any quality assurance system and at the heart of the quality process sits the seventh standard, the need to collect and analyse reliable data for decision-making and to identify what is working well and what requires further attention. The European guidance recommends evaluation of the following:

- content of the programme in the light of the latest research in the given discipline, thus ensuring that the programme is up to date;
- changing needs of society;
- students' workload, progression and completion;
- effectiveness of procedures for assessment of students;

- students' expectations, needs and satisfaction in relation to the programme;
- learning environment and support services and their fitness for purpose for the programme.

This means that effective processes to collect and analyse information about courses and qualifications need to feed into an internal quality assurance system. The following are typical key performance indicators:

- profile of the student population;
- student progression, success and dropout rates;
- students' satisfaction with their programmes;
- learning resources and student support available;
- career paths of graduates.

Other quality guidelines from Australia and the UK (QAA 2023) exhibit similar principles, noting that the information gathered by individual institutions for external appraisal and self-regulation in these countries depends, to some extent, on the type and mission of the institution. Australia, however, includes research and research training (see the Australian Government's Tertiary Education Quality and Standards Agency; TEQSA 2021). Canada does not have a national university accreditation system. Instead, all education is regulated provincially but universities tend to belong to Universities Canada, which establishes standards of quality for all Canadian degree programmes (Universities Canada 2023).

International microcredential quality standards

It is clear from the final report of the Micro-credentials Higher Education Consultation Group (European Commission 2020) that the rationale for an European approach for microcredentials

is to increase personalised learning for all and widen learning opportunities both in higher education and vocational education and training (VET) establishments. The aspiration is to mainstream microcredentials' use with respect to both an economic and social mission perspective.

It was agreed that a standards framework for microcredentials should align with national (NQF) and European (EQF) qualifications frameworks, that Member States could consider adapting their own national qualifications frameworks to include microcredentials, and that an important step in this process was to agree a transparent definition, which is:

A micro-credential is a proof of the learning outcomes that a learner has acquired following a short learning experience. These learning outcomes have been assessed against transparent standards. (European Commission 2020)

In summary, the European recommendations for a microcredential quality framework in 2020 were that it should include:

- a defined list of critical information elements to describe microcredentials;
- alignment with national qualifications frameworks (NQFs) and the European Qualifications Framework (EQF): defined levels, standards for describing learning outcomes;
- quality assurance standards;
- defined credits: European Credit Transfer and Accumulation System (ECTS), defined learning outcomes and notional workload;
- recognition: for further studies and/or employment purposes;

- portability: issuing, storage and sharing of micro-credentials;
- platform solutions for the provision and promotion of courses leading to microcredentials;
- incentives to stimulate the uptake of microcredentials.

A very important consideration that allowed the notion of microcredentials to progress was that all HEIs following ESG (Standards and Guidelines for Quality Assurance in the European Higher Education Area) quality assurance procedures could be regarded as ‘trusted providers of micro-credentials’. Additionally, where microcredentials are delivered online, the tool developed by the E-xcellence project (Rosewell et al. 2017) may be used as a reference point. There was clear recognition that microcredentials are also issued by non-higher education providers, that quality assurance is essential and that the ESG could, in principle, be used in these circumstances.

One example of how this framework has been applied is provided by the Netherlands through an ‘Acceleration Plan’ (2022). Within this plan, 32 higher education institutions (10 universities and 22 universities of applied sciences) have been taking part in a national microcredentials pilot under the direction of the Making Education More Flexible zone. The Universities of the Netherlands and the Association of Universities of Applied Sciences of the Netherlands have produced a quality framework, a starting point which can be refined and which is open to further interpretation as the universities work through it together.

1. The guideline for microcredentials is that these are educational units that are no smaller than 3 EC and no larger than 30 EC [one European Credit (EC) represents 28 study hours].

2. The education certified by a microcredential is substantively related to the institution's education and/or research portfolio. This may be existing education as well as newly developed education or research.
3. It is clear who the intended target group of the education is, what prior knowledge is required from the participants, what the entry requirements are (if any), and how these are tested.
4. The educational programme, the educational environment and the quality of the team of teachers enable the incoming participants to achieve the intended learning outcomes.
5. The learning outcomes and the educational level and scope of the microcredential are made clear. The participating institutions describe this in an unambiguous manner, in line with European agreements (Bologna) and developments in Brussels.
6. In principle, institutions recognise the (validated) learning outcomes of microcredentials that have already been attained and/or are being attained elsewhere. Whether this leads to intake and/or exemption remains within the mandate of the examination board or another body designated by the institution.
7. The tests support the learning process of the participant and the assessment is valid, reliable, transparent for participants and sufficiently independent. (Acceleration Plan 2022)

A set of minimum requirements for internal quality assurance is guaranteed by the ESG and microcredentials are to be offered in line with the lifelong vision of the awarding institution. This is similar to HEIs providing a policy for production, presentation

and assessment of any of their microcredentials advertised to potential students.

There is a consensus building around quality standards for microcredentials that should be adhered to by any recognised body which has received a quality kitemark. However, there are also other considerations such as the stackability of these credits towards a diploma or degree, and whether a transcript of these credits should be available to future employers.

A study was undertaken by the Higher Education Quality Council of Ontario (HEQCO), an agency of the Government of Ontario that undertakes evidence-based research to assist with the improvement of post-secondary education in the province. As part of its microcredential awareness investigation, HEQCO surveyed 201 Canadian employers, 161 representatives from 105 Canadian post-secondary institutions and 2,000 prospective students (Pichette et al. 2021). Their findings included the following suite of ‘quality markers’

- **Relevant:** consulted or involved industry/community;
- **Accredited:** recognised or issued by a professional accrediting body;
- **Standardised:** meets a government-set quality standard;
- **Assessed:** learner must demonstrate skills/knowledge to earn the credential;
- **Flexible:** pace and/or structure of learning can be personalised;
- **Stackable:** can be ‘stacked’ or combined toward a larger credential, e.g. a diploma or degree. (Pichette et al. 2021: 16)

These quality markers were viewed favourably by all the stakeholders and provide kitemarks that match other recommendations

from international bodies, in a clear and concise manner that would ease transferability of these types of credentials between institutions (Bates 2021), supporting lifelong learning.

National quality standards

In the UK, the Quality Assurance Agency for Higher Education (QAA) has produced a characteristics statement for microcredentials that means a set of general guidance is provided for higher education providers when developing a new provision such as microcredentials. It describes ‘outcomes and attributes of microcredentials in a UK-wide context, many higher education providers will use them as an enhancement tool for the design and approval of short courses, and for subsequent monitoring and review’ (QAA 2022).

Important considerations for the UK context include advice about how to manage the implementation of standards (which, in essence, follow EU standards) and, more importantly, the implementation and evaluation of a quality enhancement process. The QAA (2022) highlights the following areas for careful thought:

- admissions decisions, and the role of recognition of prior learning;
- approaches to course design and approval that are agile and not overly burdensome while still being robust;
- swift confirmation of outcome and award following completion of assessment;
- effective monitoring and review;
- student engagement in quality management.

Before moving on to questions of managing and evaluating the quality of microcredentials, as required by these quality standards,

it makes sense to zoom out of this level of detail and look at the context of some of these quality enhancement frameworks.

Contexts

A European approach to microcredentials will allow higher education institutions to offer such courses on a larger scale and in a comparable manner throughout Europe, ensuring agreed quality standards, and facilitating their recognition and portability across the EU (European Commission 2020: 4).

The definition of microcredentials produced by Europe's MICROBOL project (MICROBOL 2020) made reference to quality assurance in line with the Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG).

A need for just-in-time training is not easily met by a course that must go through rigorous quality assurance processes to demonstrate it meets local and national standards.

Although there is a lot of work to be done in this area, there is also the political will to achieve it, as evidenced by government initiatives supporting quality assurance and standards agencies to incorporate microcredentials within their work (see, for example, QAA 2022). At the same time, online platforms are developing pathways to study that do not necessarily lead to academic credit but do lead to industry-relevant certification.

Microcredentials need to:

- be aligned with multiple existing frameworks as well as across countries and continents;
- strike a balance between requirements for high-quality just-in-time training and the time required to carry out quality assurance processes.

The final role identified in the microcredentials project team in Chapter 3 was quality enhancement. Developing microcredentials is a large-scale strategic initiative for any institution, and including work on evaluation and quality enhancement provides opportunities to assess progress and adjust ambitions. One approach is to align aims with key performance indicators (KPIs) so that progress towards an aim such as ‘attract more international learners’ can be linked with specific targets such as: ‘1,000 registrations from countries in South America in the next calendar year’, ‘more than 50% of those who complete a microcredential successfully will be based in another country’ or ‘microcredentials offered in Mandarin will recruit as well as those offered in English’.

KPIs like these enable at-a-glance summaries of progress but quality enhancement also needs a more reflexive consideration of what has happened, what has worked well and what could be done better. Agile approaches to project management incorporate regular retrospectives, so some teams involved in microcredentials will be reflecting and developing as the initiative progresses. Other teams will have standard reporting processes that prompt them to evaluate their work and to identify opportunities for quality enhancement. An evaluation lead can bring these existing approaches together and incorporate them into a structured consideration of the initiative as a whole that can then be used by those working on the project to improve practice.

As with any course that awards academic credit at higher education level, microcredentials must be aligned with national and international frameworks. This means the normal quality assurance checks must be applied or adapted to fit them. The institution will need to be able to assure both learners and regulatory bodies that microcredentials are as rigorously checked as any other credit-bearing course and that their standards are in line with those applied to other academic courses. In Europe, for example,

all courses offered by higher education institutions must undergo internal quality assurance by the institution in question. In addition, either each course or the higher education institution as a whole is required to undergo periodic external quality assurance (e.g. accreditation, audit, review). (European Commission 2020: 14)

Processes are needed to demonstrate that a microcredential credit requires a similar amount of work at a similar standard to those qualifications on offer within the institution and more widely.

Ways of doing this will vary between institutions but might include, for example, external reviewers on microcredentials, reputable external examiners, second or third marking, scrutiny by academic committee, and agreed policies for assessment and awards that apply specifically to microcredentials. The more robust these methods are, the more helpful they will be for the credit-transfer process, which is one of the outward-facing aspects of the microcredential initiative.

A 2020 study of microcredentialing research and pilots across Canada, with a focus on their utility for admission and transfer into higher education, noted that:

If a micro-credential is to be considered as a bona fide credential ... expectations typically exist that the learning experiences (including those represented by micro-credentials) have been structured, delivered, and assessed by trusted entities in accordance with accepted and recognized quality assurance expectations and frameworks. (Duklas 2020: 15)

A part of this quality assurance is evaluation, which the Hewlett Foundation's Evaluation Principles and Practices (Twersky & Lindblom 2012: 3) define as a 'systematic' approach, stating that 'evaluation is an independent, systematic investigation into how, why and to what extent objectives or goals are achieved'.

Evaluation

Defining evaluation

However, evaluation can be a contentious issue and not all definitions agree on its purpose Gullickson (2020). Therefore, for any type of evaluation, agreement about what the term means is an essential first step to scoping the work and resources involved, together with an appraisal that considers whether the evaluation has met its objectives. It can be useful to examine some definitions of evaluation before coming to a final decision as these can prompt reflection about what can be achieved and then acted upon once the evaluation has taken place.

While the definition from the Hewlett Foundation that appears at the end of the previous section emphasises systematic elements, evaluation can be seen as a judgement of value and worth. Scriven (1991: 53) states that ‘evaluation is the process of determining merit, worth, or significance’. The findings can also aid reflection and point to future improvements. This definition could well apply to an educational pilot study, where the findings lead to a stop/go decision.

Another definition has arisen in which the essential feature is not one of judgement but of learning. Two major philanthropic bodies that fund education-related projects support the latter. The W.K. Kellogg Foundation (2017: 1) defines evaluation as a process of ‘systematically generating knowledge that can support learning, quality improvement and good judgement in decision-making’, adding that ‘evaluation also can align purpose, action and impact to ensure that longer-term change at the societal level unfolds progressively’. In its *Evaluation Handbook*, the foundation

(2012: 2) suggests that evaluation should ‘strengthen projects during their lifecycle’ and, whenever possible, provide ‘outcome data to assess the extent of change’.

Attwell (2017), focusing specifically on the evaluation of online learning, reiterates the notion of evaluation as a learning process, defining it as:

- a joint learning process for all involved, generating useful and relevant information and knowledge;
- a theoretical and practical approach, which feeds back into ongoing change processes in organisations and projects;
- a systematic process to assess the relevance, efficiency and effectiveness of policies, projects and programmes.

Evaluating online teaching

Online teaching and learning have broken new ground in that they have explicitly introduced new pedagogies and technologies, the impact of which has been evaluated and shared. These findings have influenced which new technologies or large-scale implementations receive funding or support at institutional or even national level.

There is a wealth of data that can be used to evaluate online teaching. This includes:

- learning analytics findings about students’ and educators’ use of online platforms such as virtual learning environments (VLEs);
- recordings or records of students’ discussions;
- observations of online teaching sessions.

Online teaching also has its evaluation challenges. The context in which a student is engaging with online teaching will not be apparent to the evaluator but could be of considerable significance in helping to understand the learner's experiences, attitudes, behaviour and study performance. Also, some evaluation methods, such as focus groups, can be more challenging to conduct online.

Online teaching evaluations have much in common with evaluations of face-to-face teaching. However, some aspects of online teaching and learning require specialist knowledge, such as the accessibility of online resources for students with sight or hearing difficulties. A holistic approach to evaluation is a good way forward:

A holistic assessment goes beyond course design; it acknowledges the nuances that make a course unique, including input and contributions from students, developments in the field of study, and current events. Most valuable are students' perceptions of their learning and of the course experience. A good course assessment considers the course over a period of time, and considers interactions between instructor and students, students and students, all of which create artifacts that can be studied and analyzed (Thompson, 2005). Artifacts might include, emails or forum posts of student questions, dialogue within forums, feedback from group interaction, end-of-course student surveys, LMS reports on student interaction patterns, student assignment results, and more. Course artifacts give valuable clues to a course's quality, more so when collected from two or more course iterations and analyzed collectively. (Morrison 2015)

Types of evaluation

There are various types of evaluation and a clear point of differentiation is whether an evaluation is **discrete** or **ongoing**. A discrete

evaluation has a clear beginning and end to its timeline. An ongoing example would be the collection of students' performance metrics such as assessment grades or VLE activity. The type of evaluation can also be classified in terms of the people conducting the evaluation. These might be internal staff, external examiners or a team of staff allocated to a particular project, including statistical analysts.

Different types of evaluation may take place while any course is being run or any new initiatives are being introduced. These include:

- **Performance evaluation** reports on progress towards intended goals, identifies problems and assesses whether an initiative and the resources it uses are being managed well.
- **Process (or formative) evaluations** probe the nature and quality of the implementation of an initiative. Formative evaluations are conducted during a project and identify its strengths and weaknesses. The results will typically be used to instantiate change and development and will often be carried out internally by a member of the project or course team.
- **Summative evaluations** take place after the event. These include:
 - Outcome evaluations, which aim to establish how well an initiative or programme is working overall, rather than being the basis for immediate action. Oliver (2000: 5) notes that summative evaluation 'is often an external process concerned with judgement rather than improvement', though some outcome evaluations will inform the development of an initiative before its next iteration.

- Impact evaluations, which consider what has happened as a result of an initiative. Intended and unintended impacts are analysed together with how change was achieved.

Why evaluate?

Educational evaluation can have value both in terms of the results of an evaluation and the process itself. Although Tyler's definition of educational evaluation is still used and described as 'the process of determining to what extent the educational objectives are actually being realized' (Tyler 1950: 69), Ramsden's (2003: 209) definition makes explicit that all good teaching involves not only reflection but also the evaluation of practice. He states that 'evaluation is an analytical process that is intrinsic to good teaching.' There are also regulatory licensing authorities that require evidence from evaluation that good teaching and learning are both taking place. These add another reason for sound evaluations to be conducted in agreement with national regulatory frameworks and standards.

Approaches to evaluation

First steps in planning an evaluation

A number of key planning considerations should be considered in conjunction with the evaluation focus and units of analysis. One way of refining the focus of an evaluation topic is through identifying its unit of analysis – the entity that is being analysed in the evaluation. In this way, every aspect of the evaluation will be open to inspection, allowing decisions to be made about the evaluation timing, the people involved, stakeholders,

cost, evaluation criteria, data collection approach, methodology and methods and the ethical considerations that will need to be managed.

The Kellogg Foundation (2017: 53) identifies six possible units of analysis for its funded projects. These categories can be applied to any online teaching-related evaluation, such as a microcredential course, and the descriptors would be comparable to those found below.

- **Individuals.** The evaluation focuses on the changes that individuals experience. These individuals could be microcredential learners, educators or other stakeholders such as national education policymakers.
- **Course, programme or educational initiative.** In this case, the focus is to understand whether the microcredential course or initiative is effective. This means identifying what does and does not work, together with the knowledge and skills required for educators to deliver the course or initiative and/or how the course or initiative could be improved. A formative evaluation may be particularly suitable in this instance.
- **Organisation** (for example, an entire higher education establishment). The evaluation focus could investigate changes within an organisation's priorities, culture, policies and institution-wide practices such as the introduction of microcredentials into the curriculum.
- **System** (for example, one for submitting assessments online). The evaluation will be based on a clear idea of the parts of the system that are being assessed and any changes in outcome that are to be expected. For example, where an assessment submission system has been changed to an online format, the evaluation might

consider how this has affected assessment submission rates and pass rates.

- **Policy** (for example, the nationwide introduction of a particular approach to microcredentials).
- **Community** (for example, a network of tutors delivering a new microcredential and supporting each other using social media). With this type of investigation care should be taken to clearly define the nature of the community in focus.

An evaluation can focus on one or more of these units of analysis at the same time. The size and scope of the evaluation will be informed by a number of considerations but should not lose sight of its originally funded objectives. Twersky and Lindblom (2012: 16) warn that evaluations should ‘NOT sacrifice relevance by having evaluation findings be delivered too late to matter’. Good planning and keeping deliverables to schedule are essential components of a successful evaluation that provides value for money.

Developing a logic model

‘A logic model is a graphic display or map of the relationship between a programme’s resources, activities and intended results, which also identifies the programme’s underlying theory and assumptions’ (Kaplan & Garrett 2005). It acts as a road map that represents the relationships between all the components of the model, which are usually: resources, activities, outputs and outcomes.

Logic models visually explain a project’s purpose, strategy and expected results. They help to provide clarity and identify cause and effect, including available resources to build a good plan of

work, supporting the adaptability of a project's resources and overall planning. Effective logic models make an explicit, often visual, statement of the activities that will bring about change and the expected results for the community and its people. They keep participants moving in the same direction by providing a common language and point of reference.

A logic model should convey its information on a single page. An example of a logic model that was produced by the W.K. Kellogg Foundation (1998) has five elements, represented by a line of coloured boxes arranged in a horizontal line. These five elements cover both planned work (resources/inputs and activities) and intended results (outputs, outcomes and impact). The chain of reasoning behind the ordering of these elements is:

1. **Resources and impacts.** Certain resources are needed to operate your programme.
2. **Activities.** If you have access to these resources and inputs, then you can use them to accomplish your planned activities.
3. **Outputs.** If you accomplish your planned activities then you will, hopefully, deliver the amount of product and/or service that you intended.
4. **Outcomes.** If you accomplish your planned activities to the extent that you intended, then your participants will benefit in certain ways.
5. **Impact.** If these benefits to participants are achieved, then certain changes in organisations, communities or systems might be expected to occur. (Kellogg 1998)

It is important to note that '[l]ogic models are not evaluation tools; they are learning and management tools that should be used throughout the life of a strategy, initiative or program. A logic

modelling process should facilitate effective planning, implementation, evaluation and improvement of your effort' (Kellogg Foundation 2017: 113).

The question you might ask is: 'If logic models are not evaluation tools, as they look more like management instruments, why are they important to designing an evaluation?' The answer is that the process of creating a logic model is considered to be valuable as it requires programmes to fully and clearly articulate both vision and aims, thus introducing a more structured approach to evaluation, setting out a clear hypothesis to be tested.

A specific example of a logic model is given below. It was prepared by Perryman (2021) for one of The Open University's microcredentials, *Online Teaching: Evaluating and Improving Courses*. It splits outcomes into two – short-term and intermediate – and contains an additional section in which possible evaluation questions for the course have been derived from the logic model. This worked example demonstrates how, for the purpose of evaluations, logic models give a basis for understanding how a particular programme or initiative works and its impact. This comprehension can inform all stages of the evaluation process, including the design, development of evaluation criteria and questions, data collection methods and data interpretation.

Logic model for the *Online Teaching: Evaluating and Improving Courses* microcredential.

- **Inputs:** human resources, financial resources, organisational systems, ICT [information and communications technology] and AV [audio-visual] equipment, external platforms and staff.
- **Activities:** producing the course and AV, presenting the course, recruiting and registering learners, facilitating the course, managing the assessment process.

- **Outputs:** number of learners on the course, number of learners completing the course, number of learners passing the course, number of comments in discussions.
- **Short-term outcomes:** participants achieve the course learning outcomes and gain knowledge, understanding and skills related to evaluating online teaching.
- **Intermediate outcomes:** participants conduct/plan evaluations in their own institutions, online teaching is improved on the basis of evaluation findings.
- **Long-term impact:** students at course participants' institutions benefit from improved online teaching, these students' study outcomes improve, these students' life chances improve. (Perryman, 2021)

Evaluation questions derived from the logic model.

- **Inputs:** Were the inputs sufficient and timely?
- **Activities:** Was the course developed as planned? How was the course promoted? Did the course recruit the target number of learners across identified categories? Were those learners registered effectively? Did the mentors facilitate the course as required? Was the assessment process carried out according to the required university processes and procedures?
- **Outputs:** How many learners were registered on the course? How many learners completed the course? How many learners passed the course? How many comments were made in the discussion?
- **Short-term outcomes:** Did course participants achieve the course learning outcomes in terms of knowledge, understanding and skills?
- **Intermediate outcomes:** Have course participants used their skills in conducting / planning evaluations in their

own institution? Have these evaluations been effective? Is there evidence that online teaching has been improved on the basis of the findings of these evaluations?

- **Long-term impact:** Is there evidence that students at course participants' institutions have benefited from improved online learning? Have these students' study outcomes improved? Have these students' life chances improved? (Perryman, 2021)

Evaluation questions, indicators and standards

In the Perryman logic model, the evaluation questions build on salient guiding principles, which include questions around data collection, data analysis and data reporting. There are also some general evaluation criteria that can assist with devising appropriate questions. Evaluations usually address one or more of these criteria, the exceptions being exploratory or descriptive evaluations. Nonetheless, the general criteria assist with an initial phase of question development and include effectiveness, appropriateness, implementation, efficiency, equity and need. From these general criteria, questions can be formulated, such as 'how are the intended outcomes being achieved?' This would match an effectiveness evaluation.

The types of evaluation question that are developed are also related to the type of evaluation that has been chosen, these being either formative or summative. It is important, however, to keep in mind not only the type of evaluation but also its purpose, evaluation criteria and stakeholders.

Formative evaluations take place while a course or initiative is in progress. The Open University undertook formative evaluations as it developed and presented its microcredentials. The first year's

evaluation considered progress over the year in relation to the agreed aims of the project (Papathoma & Ferguson 2020); the second evaluated production methods used for microcredentials as well as the learner and educator experience (Papathoma & Ferguson, 2021), while the third considered impact, pedagogy, assessment, and the balance between theory and practical skills (Chandler, 2023).

Formative evaluation questions relate to the ‘activities’ or ‘outputs’ stage of a logic model. Some questions that are relevant include:

- How is the programme/microcredential/project being implemented? Subquestions may focus on the enquiry and registration processes: are prospective learners’ queries answered promptly? How were learner expectations managed? Was there too much content in the courses for the allocated study time?
- How appropriate are the processes compared with relevant quality standards? Subquestions could cover any of the aspects of online teaching mentioned in related standards.
- Is the programme/microcredential/project being implemented correctly? Subquestions may be asked about how the course mentor/study adviser role was performed and whether this was found to comply with the guidance provided.
- Are as many participants being reached as intended and have any related targets been met (e.g. relating to ethnicity, gender or socio-economic status)?

Summative evaluations ask questions at the end of an initiative or programme of courses. The evaluation questions for the first phase of the OU summative evaluation (Papathoma & Ferguson

2020) of its microcredentials were based upon the agreed aims/purpose of the initiative, which, in turn, determined its effectiveness and impact. These types of evaluation are often referred to as 'outcome evaluations' or 'impact evaluations.' The types of summative question that were asked by the OU included:

- How well did the project/microcredential work? Subquestions focusing on specific aspects of the initiative included: how did the innovative approach to the development and delivery of course content work? This is an important question that also relates to the cost of the resources that produced the innovative content.
- Did the project/microcredential/programme achieve, or contribute to, its intended short-term, intermediate and long-term outcomes? Subquestions probed: did the project access international markets? Which countries favoured this form of learning and why?
- For whom, in what ways and in what circumstances? These subquestions focused on new and different learner populations.
- What external factors may have contributed to, or prevented, impact and in what ways/which circumstances? Subquestions focused on the impact of the availability/non-availability of technology and the support given by the study advisers.

Indicators are specific, measurable and observable statements that provide clearer definitions of outcome statements. Indicators guide the rest of the evaluation plan including the selection of data collection methods, the design of the evaluation instruments (e.g. the survey and interview questions used), the choice of data analysis methods, and consideration of what has occurred or

changed in the evaluated initiative, leading to further questions, such as how these changes happened.

Indicators can be quantitative, such as metrics that include the number of students who submitted a particular assessment or passed a course. On the other hand, qualitative indicators focus on variables such as attitudes, perceptions and beliefs. Indicators can relate to any part of an initiative and its logic model or initial descriptions.

There are three categories of indicator. Input indicators relate to the 'inputs' or 'resources' part of the logic model. An example from this group could relate to microcredential production costs, which are checked to see if they remained within the agreed budget. Process indicators measure the activities and related outputs to ascertain whether the initiative was implemented as planned. Some evaluations only use output indicators as their process indicators. This is based on the assumption that, if the original outputs have been achieved in a satisfactory manner, it is more than likely that an initiative's activities have been correctly implemented. Other evaluations may use separate indicators for activities and outputs. Outcome indicators measure whether the initiative achieved the expected outcome and impact identified in the logic model in the short term, intermediate term and longer term. Therefore, pre and post indicators need to be measured before an initiative starts and again at the end of that initiative. If that is not possible, then an indicator probably needs to rely on self-reported data about whether the expected changes took place.

It is likely that each activity or outcome will have more than one indicator and some indicators will be more time-consuming than others to enact. Indicators that rely on observing an educator's practice will be more time-consuming to collect evidence for than those relying on self-reports of changed practice collected via a

survey. For online teaching-related evaluations, these indicators can be drawn from existing standards and benchmarks, as discussed below.

'Standards' refers to the level of performance required for specific indicators.

'Standards' can refer to an aspect of performance, or to the level of performance, or to a combination of both. The level of performance can be specified tightly or described in terms that will vary according to the context. These standards can be considered minimum levels required, or levels required to be considered 'best practice'. (Rogers, cited in Fang 2017)

A review of standards by the International Council for Open and Distance Education (ICDE) noted in 2015 that:

There are many existing schemes and models for quality assurance of open, distance, flexible and online education, including e-learning. They share many common features and many are designed to offer flexibility for institutions to adapt to suit national and institutional contexts. The most common structure encountered presents criteria for performance in aspects of institutional management, curriculum design, student support and other elements of educational provision, further subdivision into performance indicators and indications of sources of evidence. The most general categorisation of activities is Management (Institutional strategy, visions, and resourcing) Products (processes of curriculum and module development) and Services (student, and staff support, information resources etc.). Some models apply numerical scoring criteria with target performance levels others rely on more subjective assessment of performance. There are models that require performance assessment of 20–30 items others in excess of 100. (Ossiannilsson et al. 2015: 2)

An important message from this report is that the ICDE recognised that the institutional context is likely to inform the choice of standards, or quality model that will be used in an e-learning evaluation. With respect to standards for online teaching and learning, Attwell's (2006) observation continues to be true: many online learning evaluation studies focus on the technology used rather than the pedagogy and learner experience.

As discussed at the beginning of this chapter, there is an extensive selection of standards and frameworks developed for different purposes and in differing contexts, from which can be derived generic advice about variables that need to be addressed for quality assurance and quality enhancement purposes. These have been summarised by ICDE as:

- **'Multifaceted'** – systems use a multiplicity of measures for quality, and will often consider strategy, policy, infrastructure, processes, outputs and more so as to come to a well-rounded view of holistic quality.
- **'Dynamic'** – flexibility is built into systems, to accommodate for rapid changes in technology, as well as social norms. For this reason, they rarely refer to specific technological measures, and rather concentrate on the services provided to users through that technology.
- **'Mainstreamed'** – while all the quality tools surveyed aim at high-level quality improvement, this is intended to trickle down throughout the institution and be used as a tool for reflective practice by individual members of staff in their daily work.
- **'Representative'** – quality systems seek to balance the perspectives and demands of various interested stakeholders, including students, staff, enterprise, government and society at large.

- **‘Multifunctional** – most systems serve a triple function of instilling a quality culture within an institution and providing a roadmap for future improvement, as well as serving as a label of quality for outside perspectives.’ (Ossiannilsson et al. 2015: 31)

*Involving people and stakeholders towards
an equitable evaluation*

As stakeholders play an important role in any evaluation process, developing a logic model should be a collaborative process. Each stakeholder is likely to have different opinions about elements of the logic model, especially about the mechanisms of change featured in any initiative. The process of collaboratively developing a logic model therefore requires careful facilitation to avoid conflict and allow diverse voices to be heard. These are important considerations when considering educational equity.

Educational equity is realised when there is fairness and justice for all students. This means that each student is able to develop their full academic and social potential, with the requisite support. It is therefore crucial to listen to all voices in an equitable evaluation, especially to voices that might otherwise be neglected. This can be achieved through guidelines that have been produced and which embody the concept advanced by Gorski (2016b) as equity literacy. The concept of equity literacy is important because it describes the skills and attitudes that facilitate the creation of sustainable learning environments for all. Gorski (2016a) also argues for a framework that instantiates equity literacy that can overcome some of the disparities that arise from some culture-centric guidelines.

More recent work from the Equitable Evaluation Initiative (EEI) offers a set of guidelines. The EEI set out to explore, prototype and advance a new frame for evaluative thinking, a five-year initiative that started in 2019. Its framework was expanded in May 2023 and its three principles are:

Principle one

Evaluation and evaluative work should be in service of equity:

- Production, consumption, and management of evaluation and evaluative work should hold at its core a responsibility to advance progress towards equity.

Principle two

Evaluative work should be designed and implemented commensurate with the values underlying equity work:

- Multiculturally valid, and
- Oriented toward participant ownership.

Principle three

Evaluative work can and should answer critical questions about the:

- Ways in which historical and structural decisions have contributed to the condition to be addressed,
- Effect of a strategy on different populations, on the underlying systemic drivers of inequity, and
- Ways in which cultural context is tangled up in both the structural conditions and the change initiative itself. (Equitable Evaluation Initiative 2023)

These guiding principles are worth considering when designing any educational evaluation.

Ethical evaluation

An ethical evaluation involves standards of conduct that promote integrity, honesty and respect for all the actors involved. According to Barnett and Camfield's (2016) definition, an ethical evaluation is 'a set of principles of right conduct that is supposed to govern practitioners' behaviours'. In an educational evaluation context these correct behaviours would fall within the normative ethics domain.

There are general ethics principles that can guide an education-related research study as provided by funding bodies, such as those published by the Australian Council for International Development (ACFID 2015) and the British Council Research and Evaluation Ethics Policy (British Council n.d.). These documents highlight the following considerations:

- informed consent;
- privacy protection and confidentiality of data;
- protection of participants' rights;
- doing no harm;
- data management and storage;
- transferring data electronically and keeping data safe in transit;
- dissemination and impact of the research.

Farrow (2016) concurs with these principles and offers a comparison of three different sets of ethics guidelines from the ESRC (2015), BERA (2011) and BPS (2010) in relation to informed

consent, independence, integrity, privacy and data security, full disclosure, respect for participant autonomy and the avoidance of harm / minimisation of risk. It is important to note that the principles that guide ethical practice in online research are similar to those for other research undertaken with human subjects: respect for autonomy, justice and beneficence (Kitchin 2007).

According to Gupta (2017), autonomy refers to the notion that each individual has the right to privacy and dignity. Justice refers to the notion that all research participants should be treated fairly, equitably and decently during the research process, while beneficence requires researchers to evaluate harms or risks to their participants and to attempt to minimise these and maximise the benefits to them (Kitchin 2007). Within the context of online research, 'the risk of harm can arise with disclosure of a participant's identity or other sensitive information that could expose them to the risk of embarrassment, reputational damage, or legal prosecution' (Gupta 2017).

Evaluators also need to be aware of the ethics of disclosure, for example with regard to students' engagement in discussion forums, where subjects may indulge in 'confessional' activity and 'online disinhibition' (Joinson 1998; Suler 2004). There is also much oversharing of personal information on social media sites, calling into question the matter of informed consent with respect to such information. This means that potential harm could be done by using this type of data, an issue that should be discussed while planning such an evaluation. A good resource to use during this planning phase is that of van den Berg, Hawkins and Stame (2022) in their *Ethics for Evaluation*, which provides a theoretical framework focusing on evaluations **doing no harm, tackling bad** and **doing good**.

Evaluation models

One of the most commonly used evaluation models is that of Kirkpatrick, which was developed in the 1950s and has been upgraded to a ‘new world’ version (Kirkpatrick Partners 2020). It can be used to measure both long- and short-term impact and has been employed by Lin and Cantoni (2017) and by Goh, Wong and Ayub (2018) as a framework for evaluating MOOCs. The current features of this model include:

- **Level 1: reaction** – learners’ feelings about the learning experience; and the more recent additions:
 - Engagement – ‘The degree to which participants are actively involved in and contributing to the learning experience’;
 - Relevance – ‘The degree to which training participants will have the opportunity to use or apply what they learned in training on the job.’
- **Level 2: learning** – the increase in knowledge, skill and changes in attitudes resulting from the learning experience; and the more recent additions:
 - Confidence – the belief in being able to apply the knowledge, and
 - Commitment to applying that knowledge.
- **Level 3: behaviour** – the implementation of acquired knowledge/skills in employment/other contexts; and the more recent addition:
 - Required drivers: ‘Processes and systems that reinforce, encourage and reward performance of critical behaviors on the job.’
- **Level 4: results** – the broader impact of the training on an organisation (or, by extension, any other environment

or stakeholders, though this is not covered in Kirkpatrick's original model); and the more recent addition:

- Leading indicators: 'Short-term observations and measurements suggesting that critical behaviors are on track to create a positive impact on desired results.' (Kirkpatrick Partners 2020)

Perryman (2020: 15) finds a problem with the model with respect to addressing 'the significance of contextual factors in enabling or inhibiting impact at Levels 3 and 4' and suggests that it does not offer a particularly nuanced approach to analysing complex relationships between cause and effect, or to capturing and understanding the impact of context on learners' experiences, and on changes in their attitudes and behaviour.

Kalz et al. (2015) developed a specific model survey instrument from the MOOCKnowledge Project. This project was an initiative of the European Commission's Institute of Prospective Technological Studies (IPTs), which aimed to provide data on participants studying MOOCs, in order to evaluate the impact of different groups studying MOOCs within a European context. Since the MOOCKnowledge model identifies variables that may explain the impact of a project or course on different groups of learners who have followed an identical course, it can assist course designers and providers to assess variables that could affect long-term impact goals.

Theory of change

The theory of change (ToC) was derived from the field of theory-driven evaluation (Chen 1990; Coryn et al. 2011) and then popularised by Weiss (1995). Its purpose is to make explicit underlying assumptions associated with a given initiative. This

allows the evaluation team to understand the goals and intentions of the project's designers. Reinholz and Andrews (2020) state that the benefits of ToC come through 'making the underlying rationale of an initiative explicit, it can be interrogated, assessed and revised systematically as it is being implemented'.

The ToC is often produced as a diagram that illustrates the interventions that will be applied to achieve the preconditions and long-term outcomes for the project. In this way the activities are clearly articulated and the diagram will assist with choosing interventions in a systematic and rigorous manner. Furthermore, it can demonstrate how an intervention has contributed to a chain of results that produced the intended or actual impacts. A ToC is therefore more complex than the logic model described earlier, with a more detailed exploration of the relationships embodied within the logic model. It also offers a framework to investigate cause and effect and to compare change mechanisms from various and diverse contexts.

A practical example of applying the ToC to an educational technology initiative is that undertaken by Perryman (2020). She employed a theory-of-change-based evaluation to the massive open online course on technology-enhanced learning (TEL MOOC) produced by the Commonwealth of Learning and Athabasca University. The evaluation report reveals extensive short-, medium- and long-term impact on TEL implementation and open educational practices across 32 countries. It also draws attention to the variables that limited the impact of this initiative, such as infrastructure problems and institution-related, cultural and technological barriers.

A theory of change approach was adopted as the basis for the evaluation due to its affordances in offering a systematic framework for investigating the complex

relationship between cause and effect that must be unravelled when conducting a long-term impact study, and for investigating the mechanisms of change in very diverse contexts. (Perryman 2020)

Since it is difficult to prove attribution for many interventions, Perryman (2020) in her evaluation adopts a contribution analysis (see Mayne 2012). Contribution analysis is a methodology used to identify the multiple factors that could be responsible for the short-, medium- and long-term impact of a given intervention such as a course or educational initiative. Contribution analysis does not conclusively prove an intervention has contributed to a change or set of changes. Instead, its prime function is to reduce uncertainty (Mayne 2008: 1).

The TEL MOOC ToC used three clusters of possible contributory factors (A, B and C), which were identified from the existing literature:

- **Cluster A** – making a potential contribution to the impact of TEL MOOC on participants in terms of changes in their attitudes and behaviour,
- **Cluster B** (identical to Cluster A) – making a potential contribution to the impact on TEL MOOC participants' colleagues' attitudes and behaviour, and on institutional/ policy change, and
- **Cluster C** – making a potential contribution to the longer-term impact on stakeholders other than the course participants and their colleagues (Perryman 2020: 42).

The benefit of this approach of identifying possible contributory factors in advance of the evaluation was that these were taken into account in the design of the survey and interview questions. This

ensured that analysis would not neglect these variables, giving more confidence to findings from this study.

Dissemination

There are multiple dissemination routes for any evaluation findings. The primary one for funders is usually a report but this may be accompanied by academic articles, conference presentations, media stories or outputs on social media. There are important factors to consider when sharing findings, including ensuring the report title can be easily found by search engines (a numerical internal report title will not be picked up by a search engine). The evaluation website can be designed to maximise the potential for visitors by using search engine optimisation (SEO) techniques. These can be complex but registering the site using keywords is a good start.

All these factors are important when using an institution's own institutional repository. The Open University supports Open Research Online (ORO), a platform in which its academics are required to deposit their research publications. Although a project may have a website, this may only exist for the duration of the research and using an institutional repository will guarantee longer term access to associated work. Peers can also be encouraged to publicise findings to their various networks.

Writing a blog and creating an account on a platform such as LinkedIn, BlueSky, Threads, Mastodon or Twitter/X will not only publicise the project but could also make the prospective audience aware of the evaluation methodology and findings. These types of message will also form an alert to the release of the final report. Other social media outlets, such as Instagram, facilitate the creation of an image-based narrative for the evaluation.

Facebook can be useful for community engagement, while tools such as Snapchat and TikTok also suit the needs of specific audiences and age groups.

Findings can also be shared as data via open data repositories, such as Figshare. This enables other users to combine the project data with their own and also with other data sets. This increases the impact of the evaluation. Events can be arranged around the release of the report, which will in turn increase social media exposure and generate more interest in the evaluation findings.

Dissemination activities can be fun and a range of unusual dissemination formats have been used. For example, the OU organises annual ‘Bake your PhD’ competitions, in which doctoral researchers bake goods that visually represent their research (these can be viewed on X using the hashtag #BakeYourPhD). By 2024, the annual Dance Your PhD organised by the American Association for the Advancement of Science together with the Science journal had run 16 times, with the entrants available to view on YouTube.

Further thoughts

Having discussed the role of policies and standards to ensure quality, together with checking the quality through different types of evaluations, several questions remain about how microcredentials are viewed and embedded within a tertiary education ecosystem. McGreal and Olcott (2022) suggest that the advent of microcredentials provides an opportunity for a strategic reset. There are, of course, risks to any new venture and microcredential creation can involve the breaking up and repurposing of previous parts of the curriculum. Brown and Nic Giolla Mhichíl (2021: 3) argue that microcredentials can be a ‘wolf in sheep’s clothing’ and strongly

suggest that an important starting point for any institution is to agree a strategic microcredential institutional framework. The development of this needs to be a high priority for the senior leadership team with due diligence around costs and market share of potential students that match government policies around skill gaps that can be addressed through microcredential production and presentation. Brown et al. (2023) offer practical advice for constructing and implementing a strategic framework with an examination of the business models that could be adopted. There is still a way to go but microcredentials offer learners a way to engage with new employment opportunities, especially with the fast growth of AI in the workplace, but only if the price is right. The final chapter of the book looks to the future and examines what may lie ahead for microcredentials.

References

- Acceleration Plan. (2022). *Microcredentials pilot*. Acceleration Plan: Educational Innovation with ICT. Available at <https://www.versnellingsplan.nl/en/Kennisbank/pilot-microcredentials>
- ACFID. (2015, February). *Guidelines for ethical research and evaluation in development*. Australian Council for International Development.
- Attwell, G. (2006). *Evaluating e-learning: A guide to the evaluation of e-learning*. Evaluate Europe Handbook Series Volume 2. Available at http://pontydysgu.org/wp-content/uploads/2007/11/eva_europe_vol2_prefinal.pdf
- Attwell, G. (2017). *Evaluating eLearning*. Available at <https://www.slideshare.net/GrahamAttwell/evaluating-elearning-104938005>
- Australian Government. (2021, November). *National microcredentials framework*. Australian Government, Department of Education, Skills and Employment.

- Bates, T. (2021). *Understanding microcredentials: A report from HEQCO* (31 May). Available at <https://www.tonybates.ca/2021/05/31/understanding-microcredentials-a-report-from-heqco>
- Barnett, C. and Camfield, L. (2016). Ethics in evaluation. *Journal of Development Effectiveness*, 8(4): 528–524. DOI: <https://doi.org/10.1080/19439342.2016.1244554>
- Brasher, A., Whitelock, D., Holmes, W., Pozzi, F., Persico, D., Manganello, F., Passarelli, M. and Sangrà, A. (2022). Comparing the comparators: How should the quality of education offered by online universities be evaluated? *European Journal of Education*, 57(2): 306–324. DOI: <http://doi.org/10.1111/ejed.12497>
- British Council. (n.d.). *British Council research and ethics evaluation policy*. British Council. Available at https://opportunities-insight.britishcouncil.org/sites/siem/files/field/file/news/research_and_eval_risk_and_ethics_policy.docx
- Brown, M. and Nic Giolla Mhichíl, M. (2021). *Micro-credentials untethered: A wolf in sheep's clothing?* Dublin City University. Available at <https://irelandseducationyearbook.ie/downloads/IEYB2021/YB2021-Higher-Education-08.pdf>
- Brown, M., McGreal, R. and Peters, M. (2023). A strategic institutional response to microcredentials: Key questions for educational leaders. *Journal of Interactive Media in Education*, 2023(1).
- Chandler, K. (2023). *Microcredentials evaluation phase 3 final report (internal report)*. Milton Keynes: Open University.
- Chen, H. T. (1990). *Theory-driven evaluations*. Sage.
- Coryn, C. L., Noakes, L. A., Westine, C. D. and Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2): 199–226.
- Duklas, J. (2020). *Micro-credentials: Trends in credit transfer and credentialing*. British Columbia Council on Admissions and Transfer. Available at <https://eric.ed.gov/?id=ED610420>
- Equitable Evaluation Initiative. (2023). *Framework* (21 May). Available at <https://www.equitableeval.org/framework>
- ESG. (2015). *Standards and guidelines for quality assurance in the European Higher Education Area (ESG)*. EURASHE.

- European Commission. (2017). *European pillar of social rights*. Available at https://ec.europa.eu/info/strategy/priorities-2019-2024/economy-works-people/jobs-growth-and-investment/european-pillar-social-rights_en
- European Commission. (2020). *A European approach to micro-credentials – Output of the Micro-credentials Higher Education Consultation Group – final report*. Available at <https://education.ec.europa.eu/sites/default/files/document-library-docs/european-approach-micro-credentials-higher-education-consultation-group-output-final-report.pdf>
- Fang, M. (2017). Design a study for determining labour productivity standard in Canadian armed forces food services. In: *Proceedings of the 6th International Conference on Operations Research and Enterprise Systems (ICORES 2017)*. pp. 219–227.
- Farrow, R. (2016). A framework for the ethics of open education. *Open Praxis*, 8(2): 93–109.
- Goh, W. W., Wong, S. Y. and Ayub, E. (2018). The effectiveness of MOOC among learners based on Kirkpatrick’s model. In: *Redesigning learning for greater social impact: Taylor’s 9th Teaching and Learning Conference 2016 Proceedings*. Singapore: Springer. pp. 313–323.
- Gorski, P. (2016a). Rethinking the role of ‘culture’ in educational equity. From cultural competence to equity literacy. *Multicultural Perspectives*, 18(4): 221–226. DOI: <https://doi.org/10.1080/15210960.2016.1228344>
- Gorski, P. C. (2016b). Equity literacy: More than celebrating diversity. *Diversity in Education*, 11(1): 12–15.
- Gullickson, A. (2020). The whole elephant: Defining evaluation. *Evaluation and Program Planning*, 79: 101787. DOI: <https://doi.org/10.1016/j.evalprogplan.2020.101787>
- Gupta, S. (2017). Ethical issues in designing internet-based research: Recommendations for good practice. *Journal of Research Practice*, 13(2): D1.
- Joinson, A. N. (1998). Disinhibition and the internet. In: Gackenbach, J. *Psychology and the internet: Intrapersonal, interpersonal, and transpersonal implications*. Academic Press.

- Kalz, M., Kreijns, K., Walhout, J., Castaño-Munoz, J., Espasa, A. and Tovar, E. (2015). Setting-up a European cross-provider data collection on open online courses. *International Review of Research in Open and Distributed Learning*, 16(6): 62–77. <https://doi.org/10.19173/irrodl.v16i6.2150>
- Kaplan, S. A. and Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28(2): 167–172.
- Kellogg Foundation. (1998). *Logic model development guide*. W K Kellogg Foundation.
- Kellogg Foundation. (2012). *Evaluation handbook*. Available at <https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/wk-kellogg-foundation.pdf>
- Kellogg Foundation. (2017). *The step by step guide to evaluation*. Available at https://www.betterevaluation.org/sites/default/files/WKKF_StepByStepGuideToEvaluation_smaller.pdf
- Kirkpatrick Partners. (2020). *The Kirkpatrick model*. Available at <https://www.kirkpatrickpartners.com/the-kirkpatrick-model/>
- Kitchin, H. A. (2007). *Research ethics and the internet: Negotiating Canada's TriCouncil policy statement*. Fernwood.
- Lin, J. and Cantoni, L. (2017). Assessing the performance of a tourism MOOC using the Kirkpatrick model: A supplier's point of view. In: *Information and Communication Technologies in Tourism 2017; Proceedings of the International Conference in Rome, Italy, January 24–26, 2017*. Springer. pp. 129–142.
- Mayne, J. (2008). *Contribution analysis: An approach to exploring cause and effect*. ILAC Brief No. 16. Rome: The Institutional Learning and Change Initiative.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3): 270–280.
- McGreal, R. and Olcott, D. (2022). A strategic reset: Microcredentials for higher education leaders. *Smart Learning Environments*, 9(9): 1–23.
- MICROBOL. (2020). *Micro-credentials linked to the Bologna key commitments: Desk research report*. MICROBOL.

- Available at <https://eua.eu/downloads/publications/microbol%20desk%20research%20report.pdf>
- Morrison, D. (2015). How 'good' is your online course? Five steps to assess course quality. *Online Learning Insights* (26 May). Available at <https://onlinelearninginsights.wordpress.com/tag/how-to-evaluate-course-quality>
- NZQA. (2022). *Micro-credential approval, accreditation and listing*. New Zealand Qualifications Authority. Available at <https://www.nzqa.govt.nz/providers-partners/approval-accreditation-and-registration/micro-credentials/#heading2-0>
- Oliver, M. (2000). Evaluating online teaching and learning. *Information Service & Use* 20(2-3): 83-94. DOI: <https://doi.org/10.3233/ISU-2000-202-304>
- Orr, D., Pupinis, M. and Kirdulyte, G. (2020). *Towards a European approach to micro-credentials: A study of practices and commonalities in offering micro-credentials in European higher education*. NESET Report. Publications Office of the European Union. DOI: <https://doi.org/10.2766/7338>
- Ossiannilsson, E., Williams, K., Camilleri, A. and Brown., M. (2015). *Quality models in online and open education around the globe. State of the art and recommendations*. International Council for Open and Distance Education (ICDE). Available at <https://files.eric.ed.gov/fulltext/ED557055.pdf>
- Papathoma, T. and Ferguson, R. (2020). *Evaluation of Open University microcredentials: Final report of phase one*. Milton Keynes: The Open University.
- Papathoma, T. and Ferguson, R. (2021). *Evaluation of Open University microcredentials: Final report of phase two (internal report)*. Milton Keynes: The Open University.
- Perryman, L.-A. (2020). *TELMOOC long-term impact evaluation study*. Commonwealth of Learning. Available at <https://oasis.col.org/items/f6dd8b4c-6eb5-45e9-8cad-038bda4e1e68>
- Perryman, L.-A. (2021). *Logic model and possible evaluation questions for the production of 'Online Teaching: Evaluating and Improving Courses'*. Milton Keynes: The Open University.

- Pichette, J., Brumwell, S., Rizk, J. and Han, S. (2021). *Making sense of microcredentials*. Toronto: Higher Education Quality Council of Ontario. Available at <https://heqco.ca/pub/making-sense-of-microcredentials>
- QAA. (2022). *Characteristics statement: Micro-credentials*. Gloucester: The Quality Assurance Agency for Higher Education. Available at https://www.qaa.ac.uk/docs/qaa/quality-code/micro-credentials-characteristics-statement.pdf?sfvrsn=32bda081_4
- QAA. (2023). *The UK quality code for higher education*. The Quality Assurance Agency for Higher Education. Available at https://www.qaa.ac.uk/docs/qaa/quality-code/revised-uk-quality-code-for-higher-education.pdf?sfvrsn=4c19f781_24
- Ramsden, P. (2003). *Learning to teach in higher education*. 2nd ed. Routledge.
- Reinholz, D. L. and Andrews, T. C. (2020). Change theory and theory of change: What's the difference anyway? *International Journal of STEM Education*, 7(2). DOI: <https://doi.org/10.1186/s40594-020-0202-3>
- Rosewell, J., Kear, K., Williams, K., Rodrigo, C. and Sánchez-Elvira, A. (2017). E-xcellence methodology: Lessons learned over ten years of development and implementation. In: Ubachs, G. and Konings, L. *Proceedings of the Online, Open and Flexible Higher Education Conference: Higher education for the future: Accelerating and strengthening innovation*. Milton Keynes: The Open University. Available at <https://conference.eadtu.eu/download2399>
- Scriven, M. (1991). *Evaluation thesaurus*. 4th ed. Sage.
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and Behaviour*, 7(3): 321–326.
- TEQSA. (2021). *Higher education standards framework (threshold standards) 2021*. Australian Government, Tertiary Education Quality and Standards Agency (TEQSA).
- Thompson, K. (2005). *Constructing educational criticism of online courses: A model for implementation by practitioners*. Unpublished dissertation (PhD). University of Central Florida.

- Twersky, F. and Lindblom, K. (2012, December). *Evaluation principles and practices: An internal working paper*. The William and Flora Hewlett Foundation. Available at <http://www.hewlett.org/wp-content/uploads/2016/08/EvaluationPrinciples-FINAL.pdf>
- Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. University of Chicago Press.
- Universities Canada. (2023). *Quality assurance*. Available at <https://www.univcan.ca/universities/quality-assurance>
- van den Berg, R. D., Hawkins, P. and Stame, N. (2022). *Ethics for evaluation: Beyond 'doing no harm' to 'tackling bad' and 'doing good'*. Routledge, Taylor & Francis.
- van den Heuvel, L. and Bohm, N. L. (2023). *Exploring the reliability, time efficiency, and fairness of comparative judgement in the admission of architecture students*. European Society for Engineering Education (SEFI). DOI: <https://doi.org/10.21427/WEY8-JZ69>
- Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, 1: 65–92.