

A Guide to Basic Evaluation in Higher Education

(Why needed and how
to do it)

MARCH 2021

Stella Jones-Devitt & Liz Austen
Evaluation Research Team
SCoLPP
(Staffordshire Centre of Learning and Pedagogic
Practice)



Preface

Within the uncertain context of Covid-19, seeking to understand what works in higher education has taken on crucial significance. Evaluating interventions effectively, in order to provide the fullest opportunities for all students to succeed, is needed more than ever as we all navigate an array of ambiguities, intangibles and anxieties. This short Guide is to assist practitioners to gain confidence in using basic forms of evaluation.

In putting this together, we are indebted to the work of Professor David Parsons, who has certainly demystified evaluation for us in so many ways, and also to colleagues in both SCoLPP (Staffordshire Centre of Learning and Pedagogic Practice) and STEER (Student Engagement Evaluation and Research) at Sheffield Hallam University for their support.

We look forward to learning much more from working with colleagues involved in the enhancement themes work and hope that you find this Guide of use throughout your projects.

Stella Jones-Devitt and Liz Austen
Evaluation Research Team for QAA Scotland Enhancement Themes

Contents

1.	Why evaluate?	4
	1.1 Developing an evaluative mind-set	
	1.2 Understanding the 'counterfactual' and its importance	
2.	Key planning principles	4
	2.1 Establishing a logic chain	
	2.2 Process or impact evaluation?	
3.	Important methods	5
	3.1 Randomised controlled trial (RCT)	
	3.2 Quasi-experimental (QE) methods	
	3.3 Non-experimental (NE) methods	
	3.4 Qualitative methods (QM)	
4.	Research design fundamentals	10
	4.1 Evaluation research design principles	
	4.2 Proportionality	
5.	Methodologies exemplars (learning and teaching)	11
	5.1 Using caricature to evaluate creative and critical pedagogic thinking	
	5.2 Using game-playing as a form of effective evaluation research	
	5.3 Using Integrative Reviewing for evaluating complex phenomena	
	5.4 Using Multi Modal Visual Methods as a form of evaluative story-telling	
6.	Benchmarking and reporting	19
	6.1 OfS evaluation standards for change	
	6.2 Evidence-informed gap analysis	
7.	Further support	22
8.	Summary overview	23
9.	References	24
10.	Appendices	27
	Appendix A: Evaluation checklist	
	Appendix B: ROTUR planning tool	
	Appendix C: Further useful resources	

I. Why evaluate?

Given the context of metric-informed provision and the need to report into TEF and NSS processes, there is an increasing imperative to gather evidence concerning the impact of actions upon the student experience.

1.1 Developing an evaluative mind-set

This helps you to measure change effectively. It means that you can find out whether something you have implemented has met its objectives and you can then assess how well those objectives have been realised.

1.2 Understanding the 'counterfactual' and its importance

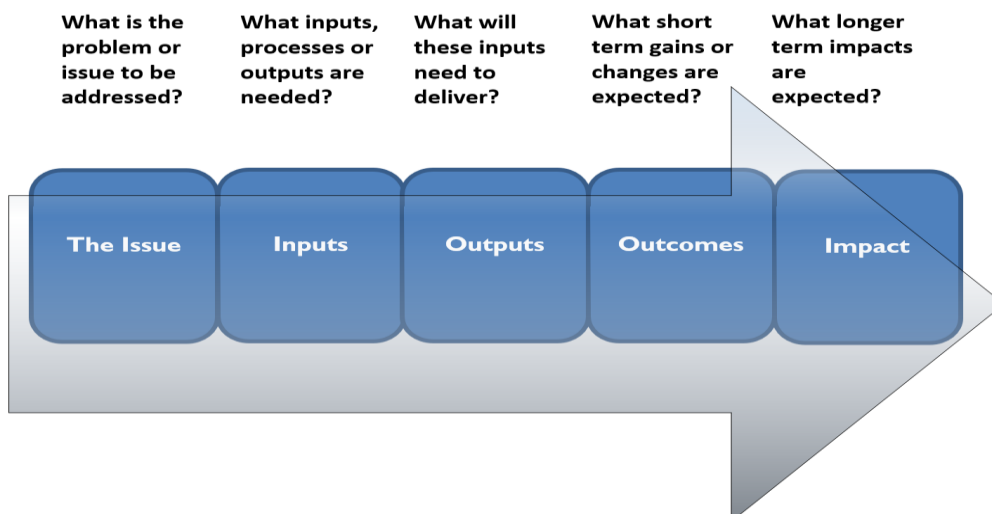
Evaluation often assesses what is known as the 'counterfactual'. This means identifying what would have occurred if an intervention or activity had not been implemented, i.e. 'business as usual' and comparing this to the measured outcomes after the intervention.

Using effective evaluation means that you can assess what works, how well it works, and what to do thereafter as a consequence

2. Key planning principles

Any evaluation starts with a basic idea, issue or problem to be addressed or explored. You then decide what inputs (such as processes and resources) are required. Inputs need to be linked to delivering some outputs. These outputs can then be related to gains (or outcomes) that you anticipate occurring as a consequence. Finally, the longer-term sustainability (or impact) of the change is considered and consolidated, revised, scaled accordingly, as detailed in the 'logic chain' diagram below (adapted from *The Magenta Book*, 2011).

2.1 Establishing a logic chain



Setting up a simple logic chain can be helpful, in which the issue, required inputs/outputs, expected outcomes and anticipated impact(s) are identified at the outset of your planning

2.2 Process or impact evaluation?

Many would-be evaluators get confused about whether they are conducting a process or impact evaluation and its importance. A **process evaluation** can be conducted whilst an intervention is ongoing to determine if it is achieving its objectives as expected. It can also be used to determine what went well, and what could have gone better. An **impact evaluation** includes assessment and measurement of the effect and influence an intervention has made. Many evaluations examine both process and impact.

Always start with evaluation when planning any kind of intervention. You need to have a baseline from which to capture and compare change (see Evaluation Checklist - Appendix A - as a basic exemplar). If you do evaluation as an afterthought you will be left guessing at some of the impact!

3. Important methods

There are lots of evaluation methods from which to choose. The first two approaches, the *Randomised Control Trial* and *Quasi Experimental*, unreservedly privilege what is known as a 'positivist' ethos (in which there is a belief that scientific research can provide an objective, and observable, understanding of any phenomenon). This has led to many quantitative analysts believing it to be a superior form of evaluation due to its underpinning empiricism:

3.1 Random Controlled Trial (or RCT)

This is used when wishing to validate the net change or impact of action(s) upon a single variable. It involves:

- Applying robust quantified analysis (statistically 'proven') of measured 'net' outcome(s)/impact(s)
- Finding differences in impact for 'intervention' and 'control' (non-intervention) participants
- Using rigorous application of probability methods (sampling/selection) in a defined population with 'matched' characteristics
- It employs demonstrable randomness in intervention/control selections
- Ensuring application of the intervention in a controlled and stable environment (minimising disturbance)
- Identifying testable timescales for securing realisable outcomes and impacts.

Within the context of higher education, use of the RCT as a viable evaluation approach of student experience is problematic:

- Setting up any kind of 'control' means greater resource needs and longer timeframes
- Randomisation requirements may be impractical in 'real-life' contexts

- Anticipated small ‘net impacts’ may require large samples which might be impractical in many H.E. contexts
- Ethical considerations may make RCTs impossible
- May only be suitable for ‘new’ and testable initiatives (not for already-established interventions)
- Not suitable for complex interventions or formative evaluation due to messiness of process and possible emergence of confounding variables

Higher education example: **Tailored support programme with control group comparison.** Comprised an individual RCT; two groups; stratified by attainment, gender, ethnicity and tested for equivalence on relevant variable. See: Bergin, D. A., Cooks, H. C. and Bergin, C. C. (2007) Effects of a College Access Program for Youth Underrepresented in Higher Education: A Randomized Experiment. *Research in Higher Education* 48 (6): 727–750.

3.2 Quasi-experimental (or QE) approach

This approach is often viewed as a more viable, practical alternative to the RCT. QE embraces some of the principles of RCTs but uses ‘comparison’ groups to take the place of actively managed control groups, meaning:

- It offers more flexibility and adaptability (beyond narrow scope and randomisation/control rigours of RCTs)
- It is suitable for measuring net impacts from more complex interventions (multiple testable outcomes)
- It is also suitable where there is limited scope for ‘control’ or when lacking sufficient data for randomising

There are some reasons to use QE with caution, as several factors make it feel like the poor relation of the RCT if looking to preserve strictly quantitative traditions, comprising:

- Any non-randomness needs robust handling of participation bias and application of appropriate principles (for example, by using ‘intention to treat’ analyses systematically)
- Some ethical constraints may be reduced but concerns around any non-intervention groups may still limit use of QE in sensitive contexts
- There are analytical challenges in identification and analysis for ‘confounding variables’ if the proposed evaluation is less linear than its RCT counterpart
- It needs to be flexible enough to analyse and/or allow for ‘unintended consequences’
- The variable robustness of different QE methods needs to be assessed by mechanisms such as the Maryland Scientific Methods Scale (SMS). The SMS is a five-point scale ranging from 1, for evaluations based on simple cross-sectional correlations, to 5 for randomised control trials

Higher education example: **Academic support, college familiarisation and career support (of Upward Bound Math-Science scheme)**. Comprised quasi-experimental design in which a retrospective comparison of Upward Bound Math-Science (UBMS) scheme participants with non-UBMS participants was done, using propensity score matching and regression analysis. See: Olsen, R., Seftor, N., Silva, T., Myers, D., DesRoches, D., and Young, J. (2007). Upward Bound Math-Science: Program Description and Interim Impact Estimates. *US Department of Education*

Positive benefits of using empirical evaluation methods are that they: provide reliable quantification of net impact; are credible and comprehensible for commissioners; evidence is more transparent and therefore more useful, and is often requested, for policy leverage.

Limitations include that they have: restricted application and scope; ethical constraints; they provide only summative not formative evaluation; they can only measure 'net' change or impact which doesn't explain the 'how or why' the change has occurred.

3.3 Non-experimental (NE) methods

NE methods are distinct from QE because they do not rely on direct comparisons to assess the counterfactual (i.e. outcomes of intervention v non-intervention groups). With NE processes, a 'comparator' is not selected, nor is it actively or passively controlled. The approach is still primarily quantifiable. This type of intervention can involve:

- The use of 'whole population' comparisons drawing on a range of evidence to assess net impacts from intervention groups
- Before and after' analyses of any change within a specific intervention group, including explanations of 'why'
- Estimations, modelling and/or trajectory analyses in any chosen intervention group(s) using pre-implementation historic data
- Possible outcome contrasts from broadly 'matched' comparator groups (which would be non-controlled or randomised)

NE methods work well when trying to establish causal relationships and associations. They are better suited to specific types of evaluation:

- When it is crucial to keep to low and/or no costs alongside having severe time constraints
- In which the impact evaluation is progressive, i.e. it uses formative as well as summative evaluation
- When it is appropriate to use methods in combination (for example, when using before and after analyses with trajectory analyses)
- NE can be harnessed with mixed mode approaches (quantitative plus qualitative data collection and analysis) to enhance scope and value
- When suitable comparator data are already available (this could be benchmark data from sector, subject area, discipline or more local source within the HEI)

There are some reasons to use NEs with caution, comprising:

- Estimation is always thought of as less reliable than measurement
- NEs lack the official credibility of RCTs and QE and are not viewed as worthy substitutes by evaluators - and often by sponsors - who give primacy to RCTs as the 'Gold Standard'
- Evaluators are faced with the challenge of isolating non-intervention influences in complex situations
- NE usually needs 'matched' comparator data
- Given the more hybridised approach, there is limited analytical confidence where 'net' impact differences are small

Higher education example: **Exploring attitudes of undergraduate learners and their lecturers towards the introduction of an attendance policy in a higher education institution.** Comprised exploring and comparing attitudes of both students and staff to the introduction of compulsory attendance monitoring. This was also cross-referenced to students' performance and progression. See: Bruen, J., Kelly, N., and Loftus, M. (2019) Attendance policies on university language degrees: exploring attitudes, and measuring relationships with performance and progression, *Journal of Further and Higher Education*. (Published online: 04 Jul 2019).

There are other more nuanced approaches to evaluation which take context into account and can attempt to provide explanations concerning causal or associative outcomes for change. These are predominantly qualitative methods as they examine meaning rather than seeking to measure net gain. Some of these methods will still want to examine counterfactual issues, whilst others are much more concerned with 'distance travelled' within a particular intervention.

Positive benefits of using NEs include that: they can be a cost-effective way of assessing forms of association and potential causality within interventions; they can be more ethically robust; they can use mixed methods approaches effectively; they can explore meaning as well as metrics.

Limitations include that they: often rely upon trajectory analysis and estimations which can be unacceptable in some evaluation contexts; difficult to sometimes isolate non-intervention influences.

3.4 Qualitative methods (QM)

Non-quantifiable, alternative qualitative methods (QM) and approaches represent a step-change in evaluation methodology. These approaches are primarily qualitative, and they seek to examine and capture meaning, connections and unintended outcomes. There are no comparators involved and outcomes tend to consider potential for scale, within a defined context, rather than direct reproducibility. This type of intervention can involve:

- Longitudinal review such as via the use of Appreciative Inquiry (Cooperrider and Whitney, 2000) as a basis for initiating, examining and potentially attributing positive change over time
- In depth qualitative case studies which can gather evidence for causal, or at least associative, impact assessment concerning what works
- Contribution analysis (Mayne, 2001) in which there is a qualitative assessment of the likely quality of attribution of impact(s) to given intervention(s)
- Qualitative Comparative Analysis (Schneider and Wagemann, 2012) is a means of analysing the causal contribution of different conditions (e. g. aspects of an intervention and the wider context) to an outcome of interest. It seeks to reduce the number of configurations needed to explain all the outcomes, a process known as minimisation
- Consideration of ‘intangible assets’ – evidence which contributes to success in HE, relating to teaching quality and student success, and which consciously goes beyond proxy indicators of student satisfaction, educational outcomes and graduate salaries. (See Robertson, Cleaver and Smart, 2019).

Positive benefits of using QMs include that: they are more easily understood by stakeholders; they can be used flexibly, affording more adaptation (especially useful for process evaluation); they can enrich understanding of context and process in ways quantitative methods cannot; they can capture unintended outcomes more effectively.

Limitations concern that: the lack of quantification hinders credibility with some stakeholders; there is greater emphasis on requisite skills levels of QM evaluation researchers; QM can be costlier and take longer to complete; selection and validation can be challenged.

Higher education example: **Evaluation of digital storytelling as an intervention for engaging in difficult conversations about positive cultural and behavioural change.** Comprised using a novel and emerging method of data collection as an evaluation tool alongside being an innovative way of sharing evidence and expertise. See: Austen, L and Jones-Devitt, S. (2018). *Observing the observers: Using digital storytelling for organisational development concerning ‘critical Whiteness’* York: AdvanceHE.

Start with the purpose and scope of the evaluation when considering methods. You need to develop a proportionate approach in any evaluation research process (see ROTUR Model in Section 4 as a basic exemplar).

4. Research design fundamentals

Effective operationalising of any evaluation scheme relies on appropriately considered design. This requires evaluation researchers to align chosen methods with achievable needs and cogent rationale.

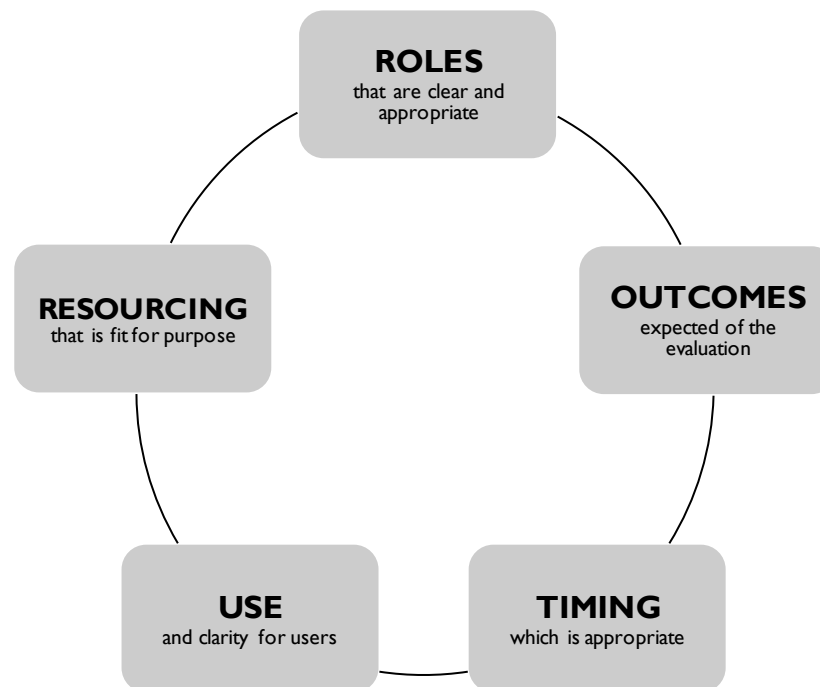
4.1 Evaluation research design principles

Parsons (2017) has identified 5 key steps for effective evaluation research design, comprising:

- Evaluation research is not incremental. It requires clarity of expectations and needs before all else
- One size does not fit all. Good evaluation research design is **always** customised to some extent
- Effective design is not just about methods. It needs to combine technical choices (contextualised, fit for purpose, robust) with political context (so it is understood, credible and practical)
- Method choices are led by a primary dichotomy: consideration of measurement (how much) vs. understanding (how; why). The chosen method can do both, if required
- Good design is always proportionate to needs, circumstance and resources

One useful source to help with effective design, amongst the many available is the **ROTUR Framework**. The ROTUR mnemonic stands for **R**oles-**O**utcomes-**T**iming-**U**se-**R**esourcing. Expectations of any of ROTUR which do not closely reflect needs are deemed ‘unrealistic’ in this process and need to be reviewed or renegotiated. Although renegotiation can feel onerous, it is worth doing this frontloading as a failed evaluation (or poor evidence emerging from it) can be more damaging. (See also Appendix B: the ROTUR planning framework).

The **ROTUR** process (Parsons, 2017)



4.2 Proportionality

Proportionality is an expected driver of design choices. The *Magenta Book* (2011) suggests that evaluations need to be proportional to the risks, scale and profile of the intended intervention, which has implications for the type and level of resources required. Key factors include:

- Stakeholders often specify that proportionality should be built into design choices at the outset
- It is necessary to consider design choices which are relative to scale and maturity of the initiative
- Resourcing constraints (principally funding / time) put pressure upon evaluation researchers to achieve 'value' through proportional choices
- Proportionality is not a negative construct. It assists evaluators to make effective design choices

There are always 'trade-offs' when applying proportionality considerations. These include whether to emphasise breadth or depth, recognising any impact on proposed level of analysis and whether trade-offs have impact on further competing needs within the evaluation process.

Regardless of which method or approach is used, all evaluation researchers need to apply rigour in planning any evaluation. Factoring in significant time to establish effective design processes and to negotiate proportionality constraints (and possible impact) is essential at the outset.

5. Methodologies' exemplars (learning and teaching)

It is really important to use a range of evaluation approaches, rather than always giving primacy to quantitative methods when assessing impact of learning and teaching approaches. The following evaluation tools and exemplars applied to pedagogic practice show how being creative can assist you in becoming a much more effective and innovative evaluator.

5.1 Using caricature to evaluate creative and critical pedagogic thinking

How Used
This activity helps to support gathering evidence concerning the art of what might be possible
Primary Purpose
To enable effective scenario-modelling/constructing of possible futures which are evidence-informed
What Happens
The process uses Romm's (1998) notion of caricature which is used as <i>a symbolic device pleading for experimentation with alternative avenues of argument, rather than as a representational device referring to 'the fact'</i> in which elements of reality are extended to their widest known limits, even to absurdity. This allows participants to play quite freely in the process by a) distancing their personal identities

so that they become part of the caricature, thus minimising risk b) liberating their thinking by considering stances they would never normally countenance, thus challenging their own assumptions but within a safe, scaffolded space. This approach can be used to generate evidence for future proofing ideas at all levels within H.E.

Getting animated about flexible learning

The development of a set of four contrasting scenarios of 'would-be' higher education institutions, and prompts, based on evidence gathered from Sheffield Hallam University stakeholders with expertise and lived experience of flexible learning (see animation). The research used Barnett's *15 Conditions of Flexibility* to frame the initial evidence-base. These scenarios have been used with a range of stakeholders to plan institutional infrastructure in several universities, related to organisational flexible learning ambitions and envisaged future positioning.

Build-a-Leader

This uses the medium of drawn caricature; in which individuals in leadership roles (can be at any level) envisage the characteristics of the most effective leader. To do this, they use a template provided, with some prompts, to evidence the qualities and attributes required. This is done individually and then this evidence is brought to a follow-up group session in which larger composite drawings are created, constructed of a meta-analysis of individual efforts.

Pedagogic Research Context

Introduces participants to evidence-informed pedagogic planning but minimises the risks as the process provides a set of possible blueprints for action, rather than evaluating things that have already occurred. Allows for lots of 'thinking the unthinkable' within a safe space. Encourages creative collective thinking and can result in many unintended consequences both as a process of engagement and evidence-gathering.

Useful Links

Jones-Devitt, S. (2016) *Getting animated about flexible learning* Downloadable project resource at: <https://www.heacademy.ac.uk/knowledge-hub/getting-animated-about-flexible-learning>

Romm, N. (1998) 'Caricaturing and Categorising in Processes of Argument' *Sociological Research Online*, 3(2). Available at: <http://www.socresonline.org.uk/3/2/10.html>

5.2 Using game-playing as a form of effective evaluation research

How Used

This activity helps to developing pedagogic opportunities that animate critical thinking processes from within a culture of experimentation and play

Primary Purpose

To facilitate problem-solving processes for anything from straightforward decision-making to examining complex subject matters in a variety of contexts. Game-playing aligns with the work of Vygotsky (1978) as a process exploring the zone of proximal development, in which the distance between learners' ability to perform tasks collectively with peer support and individuals' ability to solve problems independently is blurred and reduced, leading to increasing reciprocal autonomy

What Happens

Participants undertake structured game-playing opportunities in order to consider the value of collective critical thinking as form of gathering more informed evidence than could be achieved when working individually. Two examples of this process comprise:

So What? Snakes and Ladders Critical Thinking Game

This is Snakes and Ladders with a difference: the ladders and snakes are separated from the board and participants make critical and collective decisions about what each moveable ladder (enabler) and each moveable snake (inhibitor) represent, dependent on the context of the pedagogic scenario presented. Players conceptualise what the 1 - 100 squares represent and consider if there is significance to specific numbers. An additional layer of complexity is introduced as participants are provided with a set of task cards about a particular issue, which can be prepopulated or – ideally – generated by participants themselves in a preparatory session. Upon landing on a snake or ladder, one participant selects a task card and completes the task within a time-limited period, after which all other game participants form a consensus view to decide whether the individual has earned the right to stay or move, as judged by the quality of their response. This process has been designed to engage participants in performing collaborative pedagogical thinking rather than as pure application of analytical and cognitive proficiencies. It provides the chance to explore and capture consensus evidence about a range of pedagogical issues.

The origami frog game

Participants are given a task to complete - *the origami frog* - and assume that the primary purpose concerns visual translation of critical thinking skills (especially as the instruction sheet is primarily diagrammatic). This is not strictly true. The task is supposed to be completed within 5 minutes and it really concerns engagement and leadership approaches within a defined group when given unfamiliar information and a novel task. Experiential evidence is often prevalent (e.g. the facilitator has stated that she can do the task in the allotted time, there are worked examples of various stages of completion, others in the room may have relevant origami skills) yet participants often ignore the evidence and group resources around them, rushing instead to show individual task accomplishment and becoming frustrated when they cannot show empirical evidence of success. During the process, group performance is observed by peers and their observations about the

game are fed back. This is a great method for demonstrating how valuable unintended consequences can be missed if only focused upon empirical outputs.

Pedagogic Research Context

Introduces participants to notions of effective collaborative pedagogic research, whilst minimising risk-aversion and highlights the value of challenging assumptions about a) an over-reliance on empirical, linear evidence b) discarding unintended consequences as these might contribute to more effective pedagogical insights.

Useful Links

Itten, J. (1919) in F. Tierney (2010) Toward an eccentric (design) pedagogy. *Design Principles and Practices: An International Journal* 4 (1), 435-441.

Jones-Devitt, S. (2013) Performing critical thinking? Chapter in edited book in conjunction with Association of National Teaching Fellows, showcasing excellence in teaching in T. Bilham (Ed.) *For the Love of Learning: innovations from outstanding university teachers* Houndmills, Basingstoke: Palgrave Macmillan Ltd.

Vygotsky, L. S. (1978) *Mind in society*. Cambridge MA: Harvard University Press.

5.3 Using Integrative Reviewing for evaluating complex phenomena

How Used

This activity helps to build a body of evidence from a variety of sources and then test iteratively with identified expert stakeholders.

Primary Purpose

To enable the thorough exploration of complex phenomena using a critical appraisal of evidence. The process is grounded in a review of the existing evidence which is then contextualised and critiqued by an expert reference group. This critique is also collated and analysed as data to add to the integrative review (IR). This process may be iterative until theoretical saturation/the limits of proportionality have been reached.

What Happens

This rigorous process of analysis and synthesis follows a determined 5 stage process:

I. Problem Identification: providing clear identification of the problem, purpose of the review and variables, to provide focus and clear boundaries.

2. Literature Search: creating a well-defined and documented search strategy (search terms, databases used, varied search strategies, and inclusion and exclusion criteria) acknowledging limitations. Stage 1 and 2 would lead to the creation of a Review Protocol.
3. Data Evaluation: whilst there is no prescribed way to evaluate the quality of data sources in an IR, the approach should be guided by the types of sources used and carried out using a transparent and fully articulated quality assurance process in alignment with recognised and established protocols.
4. Data Analysis: using constant comparison methods (data reduction, display, comparison, conclusion and verification) to extract themes, patterns and relationships that form the basis of conclusions.
5. Presentation: of generated conclusions clearly linked to evidence, including explicit identification of limitations and reflections on the review process.

At each stage, an expert reference group can be consulted for appraisal and guidance. These discussions are included as data and included in the prescribed analysis.

Digital capability and teaching excellence: an integrative review exploring what infrastructure and strategies are necessary to support effective use of technology enabled learning (TEL) (QAA Subscriber Research)

In this example, an integrative review was applied to research into digital capability and teaching excellence - two fluid and contested terms. It was anticipated that good practice guidelines for developing digitally-capable teaching excellence would be established as an outcome of the project through a process of identifying and synthesising available literature and via continuous evaluation by wider expert reference groups. An 'Internal Digital Capability Steering Group' was selected, including colleagues at the host university, and an 'External Digital Capability Steering Group' was selected to utilise external expert stakeholders from across the HE/FE sector. Throughout the review, the expert steering groups provided direction for the operationalisation of the research question (defining key terms), refining search terms, guidance on appropriate databases and inclusion/exclusion criteria (e.g. national/international coverage). This data was gathered during face to face sessions (internal) and web-based interactive sessions (external). The recognition of expert opinion as valid sources of evidence, along with the experts' signposting to key artefacts, influenced all stages of the process. This richness of data and resultant outcomes would not have been achieved by relying solely on traditional forms of synthesising evidence.

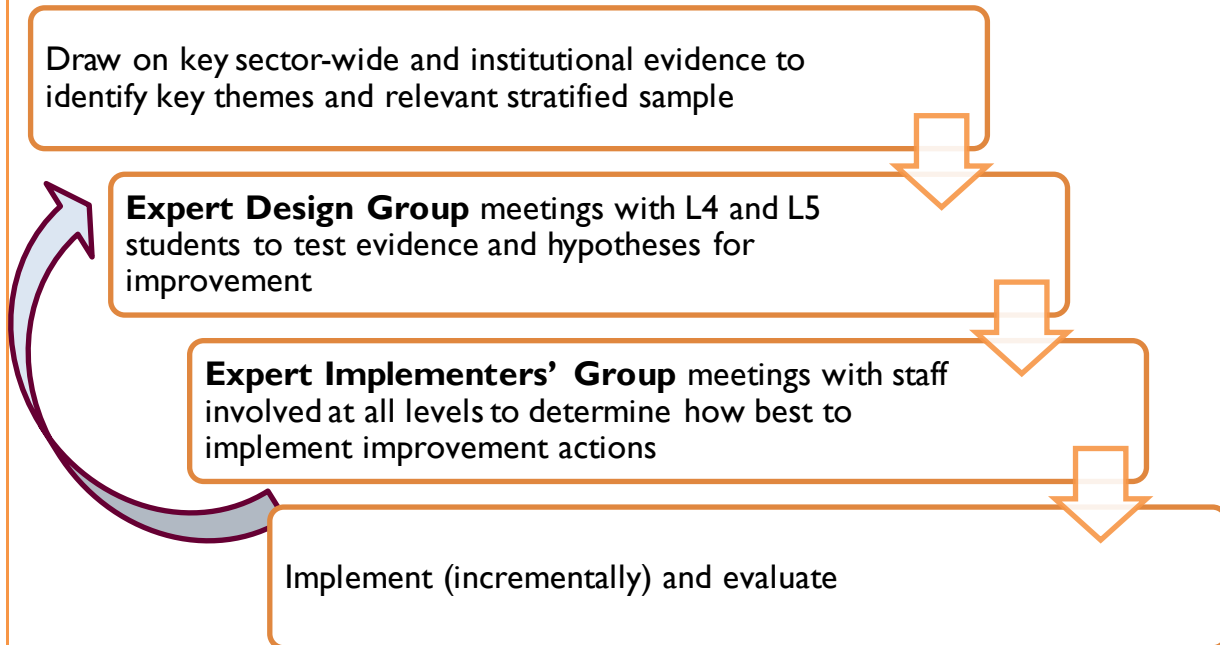
Class of 2020 - a new way of engaging students in real-time decision-making about their learning experience

This project aimed to track students throughout their undergraduate studies at Hallam by implementing student-involved research. The project used a longitudinal integrative review process in an effort to gain insight into ways of improving retention and overall student outcomes. Key principles of IR methodology were used, in which: Integrative Reviews go beyond traditional boundaries of systematic reviewing by welcoming experts as valid sources of evidence and as providers of continuous data collection and synthesis (Jones-Devitt, Austen and Parkin, 2017: 1). Within the context of Class of 2020, students are defined as having expertise of their own learning and are seen as 'expert designers' of their own experience, alongside staff, who are viewed as 'expert implementers' of that experience.

The Class of 2020 pilot consisted of three phases. These concern:

1. Gathering evidence to inform the shape of the initiative. A Route Map (also known as a Protocol in IR methodology) was constructed.
2. Testing of evidence and ideas gathered in activity driven focus groups (steering group meetings).
3. Implementation of change and evaluation.

It should be noted that whilst the first phase was a discrete activity, phases two and three overlap as the change process was intentionally iterative and cumulative, occurring from the earliest stages of identifying ideas for implementation.



Pedagogic Research Context

Integrative Reviews can go beyond traditional or systematic literature reviews on pedagogic topics and can build in expert stakeholder to help synthesise findings. Pedagogic integrative reviews could position student cohorts as experts alongside those responsible for the oversight of learning and teaching innovation within the Faculty/School/Department/Institution.

As with any pedagogic research, ethical implications require careful consideration. In this case this includes obtaining longitudinal consent (at each point of data collection) and detailed briefs and debriefs to cover confidentiality within steering group meetings. Any targeted sampling approaches or incentives, which may glean continued student engagement and a diversity of voices, also need significant ethical scrutiny.

Useful Links

Austen, L., Parkin, H. J., Jones-Devitt, S., McDonald, K., and Irwin, B. (2016). *Digital capability and teaching excellence: an integrative review exploring what infrastructure and strategies are necessary to support effective use of technology enabled learning (TEL)*, Gloucester: Quality Assurance Agency, available at <http://dera.ioe.ac.uk/27672/1/Digital-capability-and-teaching-excellence-2016-summary-report.pdf>

Jones-Devitt, S., Austen, L., and Parkin, H. J. (2017). Integrative Reviewing for exploring complex phenomena. *Social Research Update* (66), <http://sru.soc.surrey.ac.uk/SRU66.pdf>

Webinar produced by Liz Austen for QAA Scotland 'Capturing Student and Staff Voices': <https://www.enhancementthemes.ac.uk/current-enhancement-theme/optimising-existing-evidence/webinar-series>

5.4 Using Multi Modal Visual Methods as a form of evaluative story-telling

How Used

This activity helps to develop and evaluate digital reflective accounts of a wide range of experiences within H.E.

Primary Purpose

To enable the digital capture of personal stories, controlled by the storyteller. These stories are supported by a combination of text, audio recordings, images, music and animations to create short films with duration of typically two to five minutes. Digital storytelling exists in numerous different formats, from multimedia online videos to image-only stories, podcasts or blogs entries, all of which contain some form of narrative produced and shared digitally.

What Happens

Participants are trained to undertake a digital story, usually during a workshop. After some guidance, each storyteller is encouraged to free write and storyboard their ideas before choosing images, associated text and finally, recording a voice-over narrative (if needed).

Digital Storytelling Within the Curriculum. Analysed for Enhancement

In this example, students were asked to complete a digital story as part of their personal development planning (formative assessment) at the end of a semester long module. These stories aimed to capture student experiences of the first 3 months of university on an extended degree in Art and Design. With permission from each student, these stories were analysed deductively to explore whether i) digital storytelling (images, narration, production) is an effective medium for student reflection; ii) the experience of extended degree students pre HE and during transition is unique and under explored; and iii) an effective pedagogy for extended degrees needs to consider variations in learner autonomy. The meta-analysis (conducted collaboratively by two researchers) provided an overview of cohort experience and has provided evidence for course enhancement, specifically around transition pedagogy.

Digital Storytelling to Explore Organisational Development

An example of this is the *Observing the observers: Using digital storytelling for organisational development concerning 'critical Whiteness' project* which specifically tested use of digital storytelling as a conduit for debate in several ways: as an intervention for engaging in difficult conversations about positive

cultural and behavioural change; as a method of data collection; as an innovative way of sharing evidence and expertise. All research participants watched a digital story made by the researchers and discussed as part of a focus group. All participants were then supported to produce their own digital story on the topic. The focus group discussions and the participants' digital stories were treated as research data and analysed thematically. In this example, an institutional sample was recruited from across the staff population at one Higher Education Institution.

Pedagogic Research Context

Digital Stories as Data

Introduces digital data capture* using an approach which gleans an authentic participant voice. Digital stories can create qualitative data as an alternative to interviews or focus groups.

Digital Stories as Prompts

Introduces digital stories as input which facilitates difficult conversations in a focus group or interview situation.

*As with any pedagogic research, ethical implications require careful consideration. In this case this includes the use of images (copyright and permissions), anonymity and associated risk (of storyteller and anyone implicated within), and confidentiality and consent (to use as data and/or share).

Useful Links

Resources produced by Liz Austen for QAA Scotland include a "How to" Guide, an example Consent Form and an Ethical Checklist for Digital Storytelling: <https://www.enhancementthemes.ac.uk/current-enhancement-theme/optimising-existing-evidence/webinar-series>

Austen, L and Jones-Devitt, S. (2018). *Observing the observers: Using digital storytelling for organisational development concerning 'critical Whiteness'* York: AdvanceHE.

Austen, L., Jones, M., and Wawera, A. (2019) Exploring digital stories as research in higher education, *Social Research Practice*, Vol 7 <http://the-sra.org.uk/journal-social-research-practice/>

For examples of completed digital stories (staff and students), and a *Digital Practice Guide*: <https://blogs.shu.ac.uk/steer/digital-storytelling-shu/>

Contrasting qualitative methodological approaches show how you can use innovative evaluation processes to examine rich meaning in your learning and teaching practice, rather than in seeking only to measure net gain.

6. Benchmarking and reporting

Often it is important to consider not only an intervention or initiative as a contained activity, but also as something to compare and / or contrast with other similar approaches within the sector; especially those deemed to be good or best practice (although, arguably, the latter is more spurious).

The process of comparing and reporting performance against equivalent external sources is known as benchmarking. Exemplars of overt benchmarking within a higher education context include national student satisfaction surveys of various guises, such as NSS, UKES, and PTES. Providers also use tools such as staff engagement surveys to gauge insights of organisational culture and market positioning. Such benchmarking is often segmented with analyses and comparisons of relevant sub-populations to provide what is known as sufficient 'granularity', i. e. evaluations that seek to characterise the scale or level of detail in a set of data and rationalise its usefulness accordingly.

6.1 Office for Students (OfS) evaluation standards for change

There has been a marked shift in higher education to try and identify 'what works' across a range of contexts. With the emergence of [Office for Students](#) (OfS) it becomes clear that evidence-informed and evidence-driven practice are assumed standpoints from which evaluation can occur. In the [Strategy for evidence and evaluation in access and participation](#) they assert why evidence and evaluation matters to them as the regulator (emboldened text provides their emphasis):

- We want to **eliminate inequality** in higher education.
- To do this we need to be **ambitious, prioritise** and **understand** what works.
- There are good examples of providers who have applied evidence and evaluation effectively to **drive improvements. We need to see more of this in action.**
- There is **limited systematic evaluation** to learn what is and isn't working and where improvements must be made.
- We want to **work with providers** and their partnerships so that we have the **most positive impact** possible.
- Evidence and evaluation can facilitate **co-creation and engagement** which is essential for **sustaining transformational change.**
- This is our **opportunity to learn more** about how to improve the impact of our work for the benefit of students, potential students and society.

OfS has funded an evidence exchange for the HE sector, known as [TASO](#) (Transforming Access and Student Outcomes in Higher Education). TASO provides an independent hub for higher education professionals to access leading research, toolkits, evaluation techniques and more to help widen participation and improve equality within the sector. OfS also commissioned research into establishing some broad-based standards of evidence and evaluation, comprising the following typology:

	Description	Evidence Used	Possible Claims
Type 1: Narrative	The evaluation provides a narrative and coherent theory of change to motivate the selection of activities in the context of a coherent strategic approach.	Evidence of impact elsewhere and / or in the research literature on effectiveness, or from your existing evaluation results.	Can provide a coherent explanation of what and why. Claims are research-based.
Type 2: Empirical	Evaluation collects data on outcomes and impact and reports evidence that those receiving an intervention might have better results, although this does not establish any direct causal effect.	Qualitative and / or quantitative evidence of a pre / post treatment change or a treatment / non-treatment difference.	Demonstrates that interventions might be associated with potentially beneficial results.
Type 3: Causal	Evaluation methodology provides evidence of causal effect of an intervention.	Qualitative and / or quantitative evidence of a pre / post treatment change on a treated group relative to an appropriate control or comparison group, using an appropriate and robust research design process.	Intervention causes improvement and demonstrable difference, using a control or comparison group.

For many Universities, establishing the first type of evidence 'Narrative' is fairly straightforward as, to a limited extent, is the second type 'Empirical' in which associations and clear pre-post intervention benefits can be assumed BUT not proven to have a direct relationship. The move towards establishing causal links in evaluation approaches is escalating and, whenever possible, this should be pursued in a proportionate manner.

6.2 Evidence-informed gap analyses (also known as 'gap maps')

Evidence-informed 'gap maps' are thematic evidence collections covering a range of issues, usually concerning large-scale policy matters in international development, they but could have increasing relevance for Higher Education, especially in helping to assess the efficacy of large-scale policy

interventions. They aim to provide an accessible visual overview of existing systematic reviews or impact evaluations in a sector or subsector, schematically representing the types of interventions evaluated and outcomes reported.

As Snilstveit et al., (2013) note that such gap maps enable policy makers and practitioners to examine findings and quality of existing evidence and this can enable informed judgment and evidence-based decision making. The gap map also identifies key 'gaps' where there is a dearth of evidence from impact evaluations and systematic reviews; hence, this identifies where future research could be focused. Thus, gap maps can be a useful tool for developing a strategic approach to building the evidence base in a particular sector.

Key objectives of evidence gap maps (adapted from Snilstveit et al., 2013)

- To facilitate speedily-informed judgment and evidence-based decision making in policy and practice. This is done by providing user-friendly tools for accessing evidence which enable policy makers and practitioners to explore findings and quality of existing evidence concerning a topic quickly and efficiently.
- To facilitate strategic use of scarce research funding to enhance the potential for future evidence synthesis by identifying key 'gaps' in the available evidence-base indicating where future research could possibly be focused.

Evidence gap maps do have some limitations. They can be extremely useful when scoping the evidence-base for any proposed intervention, but they do not seek to answer a specific research question; instead, focusing on providing a broad overview of the existing evidence. They are also restricted to studies that assess effectiveness of interventions and therefore do not include evidence on predictive factors, implementation, barriers and facilitators to effectiveness and other types of evidence. Evidence gap maps do not provide details about the wider contextual background of the included evidence, nor do they synthesize the findings of included systematic reviews and impact evaluations. Most importantly, evidence gap maps do not aim to provide recommendations or guidelines for policy and practice, per se, but rather to be one of the crucial evidence-sources that inform policy development and guidelines.

Appropriate use of benchmarking can be very helpful for robustness of evaluation approaches and can stop interventions from becoming insular. There are many useful tools in the sector to assist in developing 'what works' principles.

7. Further support

We want to support colleagues undertaking enhancement themes project work to be in a position to confidently design and implement evaluation processes that facilitate robust evidence-informed approaches. If you would like further support in this area, please do contact Stella or Liz by email via stella.jones-devitt@staffs.ac.uk or at scolpp@staffs.ac.uk

We would be delighted to hear from you!

8. Summary overview

Using effective evaluation means that you can assess what works, how well it works, and what to do thereafter as a consequence

Setting up a simple logic chain can be helpful, in which the issue, required inputs/outputs, expected outcomes and anticipated impact(s) are identified at the outset of your planning process

Always start with evaluation when planning any kind of intervention. You need to have a baseline from which to capture and compare change (see Evaluation Checklist - Appendix A - as a basic exemplar). If you do evaluation as an afterthought you will be left guessing at some of the impact!

Begin with the purpose and scope of the evaluation when considering methods. You need to develop a proportionate approach in any evaluation research process (see ROTUR Model in Section 4 as a basic exemplar and Appendix B).

Regardless of which method or approach is used, all evaluation researchers need to apply rigour in planning any evaluation. Factoring in significant time to establish effective design processes and to negotiate proportionality constraints (and possible impact) is essential at the outset.

Contrasting qualitative methodological approaches show how you can use innovative evaluation processes to examine rich meaning in your learning and teaching practice, rather than in seeking only to measure net gain.

Appropriate use of benchmarking can be very helpful for robustness of evaluation approaches and can stop interventions from becoming insular. There are many useful tools in the sector to assist in developing 'what works' principles.

9. References

Austen, L., Parkin, H. J., Jones-Devitt, S., McDonald, K., and Irwin, B. (2016). *Digital capability and teaching excellence: an integrative review exploring what infrastructure and strategies are necessary to support effective use of technology enabled learning (TEL)*, Gloucester: Quality Assurance Agency, Available online at: <http://dera.ioe.ac.uk/27672/1/Digital-capability-and-teaching-excellence-2016-summary-report.pdf>

Austen, L and Jones-Devitt, S. (2018). *Observing the observers: Using digital storytelling for organisational development concerning 'critical Whiteness'* York: Advance HE.

Austen, L. (2018) Webinar for QAA Scotland 'Capturing Student and Staff Voices'. Available online at: <https://www.enhancementthemes.ac.uk/current-enhancement-theme/optimising-existing-evidence/webinar-series>

Austen, L (2018) Resource for QAA Scotland include a "How to" Guide, an example Consent Form and an Ethical Checklist for Digital Storytelling. Available online at: <https://www.enhancementthemes.ac.uk/current-enhancement-theme/optimising-existing-evidence/webinar-series>

Austen, L., Jones, M., and Wawera, A. (2019) Exploring digital stories as research in higher education, *Social Research Practice*, Vol 7. Available online at: <https://the-sra.org.uk/common/Uploaded%20files/Social%20Research%20Practice%20Journal/social-research-practice-journal-issue-07-winter-2019.pdf>

Bergin, D. A., Cooks, H. C. and Bergin, C. C. (2007) Effects of a College Access Program for Youth Underrepresented in Higher Education: A Randomized Experiment. *Research in Higher Education* 48(6): 727–750.

Bruen, J., Kelly, N., and Loftus, M. (2019) Attendance policies on university language degrees: exploring attitudes, and measuring relationships with performance and progression, *Journal of Further and Higher Education* 44(7): 984-997. (Published online).

Cooperrider, D. L. and Whitney, D. (2001) Exploring Appreciative Inquiry *Perspectives on business and global change* 14(2): 69-78.

H M Treasury (2011) *The magenta book: guidance for evaluation*. Available online at: <https://www.gov.uk/government/publications/the-magenta-book>

Itten, J. (1919) in F. Tierney (2010) Toward an eccentric (design) pedagogy. *Design Principles and Practices: An International Journal* 4(1): 435-441.

Jones-Devitt, S. (2013) Performing critical thinking? Chapter in edited book in conjunction with Association of National Teaching Fellows, showcasing excellence in teaching in T. Bilham (Ed.) *For the Love of Learning: innovations from outstanding university teachers* Houndmills, Basingstoke: Palgrave Macmillan Ltd.

Jones-Devitt, S. (2016) Getting animated about flexible learning. Downloadable project resource. Available online at: <https://www.advance-he.ac.uk/knowledge-hub/getting-animated-about-flexible-learning>

Jones-Devitt, S., Austen, L., and Parkin, H. J. (2017). Integrative Reviewing for exploring complex phenomena. *Social Research Update* (66). Available online at: <http://sru.soc.surrey.ac.uk/SRU66.pdf>

Maryland Scientific Methods Scale (SMS). Available online at: <https://whatworksgrowth.org/resources/the-scientific-maryland-scale/>

Mayne, J. (2001) Addressing attribution through contribution analysis: using performance measures sensibly. *Canadian Journal of Programme Evaluation* 16: 1-24.

Office for Students (2020) Strategy for evidence and evaluation in access and participation. Available online at: <https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/evaluation/strategy-for-evidence-and-evaluation-in-access-and-participation/>

Olsen, R., Seftor, N., Silva, T., Myers, D., DesRoches, D., and Young, J. (2007). *Upward Bound Math-Science: Program Description and Interim Impact Estimates*. US Department of Education. Available online at: <https://www2.ed.gov/rschstat/eval/highered/upward-math-science/complete-report.pdf>

Parsons, D. (2017) *Demystifying evaluation: practical approaches for researchers and users*. Bristol: Policy Press, University of Bristol.

Robertson, A, Cleaver, E, & Smart, F (2019). Beyond the metrics: identifying, evidencing and enhancing the less tangible assets of higher education. Available online at: <https://www.enhancementthemes.ac.uk/evidence-for-enhancement/defining-and-capturing-evidence/the-intangibles-beyond-the-metrics>

Romm, N. (1998) 'Caricaturing and Categorising in Processes of Argument' *Sociological Research Online*, 3(2). Available online at: <http://www.socresonline.org.uk/3/2/10.html>

Schneider, C. Q. and Wagemann, C. (2012) *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis (QCA)* Cambridge: Cambridge University Press.

Snilstveit, B., Vojtkova, M., Bhavsar, A. and Gaarder, M. (2013) Evidence Gap Maps: A Tool for Promoting Evidence-Informed Policy and Prioritizing Future Research, *Policy Research Working Paper 6725*, The World Bank Independent Evaluation Group Public Sector Evaluation Department Open Knowledge Repository. Available online at:

<https://openknowledge.worldbank.org/handle/10986/16941>

TASO (Transforming Access and Student Outcomes in Higher Education) Evaluation guidance and resources. Available online at: <https://taso.org.uk/evidence/evaluation/>

Vygotsky, L. S. (1978) *Mind in society*. Cambridge MA: Harvard University Press.

10. Appendices

Appendix A: Evaluating the overall impact of interventions: a ten point framework.

Appendix B: The ROTUR planning framework of analysis.

Appendix C: Useful resources

Appendix A: Evaluating the overall impact of interventions: a ten point framework

Rationale

Table 1 below outlines steps each initiative should apply at planning stage to ensure robust evaluation of impact. Key questions to address are in the left hand-column with suggested (non-compulsory) sub-theme prompts for covering further detail, as deemed appropriate to notional scale and context.

Table 1. Ten point framework for planning an evaluation of proposed activity or intervention	
Key question	Suggested sub-themes
1. Describe the subject of the initiative	<ul style="list-style-type: none"> • What is its aim? • What are the main features? • What is its context and likely future?
2. State the purpose of the evaluation	<ul style="list-style-type: none"> • What is the purpose? (For example, to assess how funding has been spent, or to inform future strategies or actions?) • How will the purpose of the evaluation influence the focus and questions that are asked?
3. Explain the focus of the evaluation	<ul style="list-style-type: none"> • What is the evaluation trying to find out or show? • Why is this important, and to whom? (Identify key stakeholders.) • What are the big questions, or hypothesis/ses, that the evaluation will try to answer?
4. Define the scope of the intended evaluation	<ul style="list-style-type: none"> • What type of things will the evaluation include within its focus? • What is feasible to cover? • Are there any specific inclusion criteria?
5. State any exclusions	<ul style="list-style-type: none"> • What will the evaluation not include and why? • What is beyond the scope of the present evaluation? • Are there any specific exclusion criteria?
6. Describe the evaluation criteria	<ul style="list-style-type: none"> • What outcomes are being captured to examine and report impact? • What indicators or measures are being employed to judge impact? • How will any unintended outcomes be treated?
7. Define the evaluation questions	<ul style="list-style-type: none"> • If the evaluation uses interviews or surveys to collect data, what questions will be asked? • Are the right questions being asked in the right format? (How do you know?)

	<ul style="list-style-type: none"> • Could you refine these questions or ask them in a different way?
8. Explain the evaluation methods	<ul style="list-style-type: none"> • What type of evaluation is being carried out? (For example, cost/benefit evaluation, Social Return on Investment, Realistic Evaluation, Appreciative Inquiry, user-led evaluation, other approaches?) • What are the main components, processes or stages of the evaluation approach? • Will any baseline data be captured to indicate a starting point for impact? • What are the specific methods used to gather data? (For example, analysis of quantitative data, interviews, focus groups, survey methods, mixed, something else?) • Who will design the methods and who will undertake them? • What is the envisaged role of researchers and other stakeholders in the process? • How will findings be reported?
9. Outline the required resources	<ul style="list-style-type: none"> • What costs, time, knowledge, skills, or other resources are needed? • How much resource are you prepared to put into the evaluation? • Where will you stop? • What is the justification for use of these resources?
10. Anticipate potential consequences	<p>Ethical issues:</p> <ul style="list-style-type: none"> • How will data be stored in accordance with GDPR guidance? • Will informed consent be obtained from those involved in the evaluation? • How will you ensure harm minimisation and appropriate confidentiality? • Are there any possible intellectual property issues? <p>Organisational issues:</p> <ul style="list-style-type: none"> • Who might be excluded or lose out because of the evaluation and / or findings? • How will any unintended organisational consequences be detected and treated? • What might be the wider institutional and / or sector response to the results?

Appendix B: The ROTUR planning framework of analysis

Proposed intervention:	
ROLES	
Planning	
Design	
Doing	
Using	
OUTCOMES	
Concise	
Credible	
Logical	
Evidenced	
Achievable	
TIMING	
Scoping	
Delivery	
Review	
Reflection	
USE	
Utility of evidence	
Primary users	
Secondary users	
Stakeholders?	
RESOURCING	
Staff time	
Budget	
Governance	
Procurement?	
Communications	
Overall timescale	

Appendix C: Further useful resources

Govt. UK *Evaluation methods*. Available online at:

<https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/evaluation-methods>

Govt UK What works centres. Available online at: <https://www.gov.uk/guidance/what-works-network>

OECD *Enhancing Research Performance through Evaluation, Impact Assessment and Priority Setting*.

Available online at: <https://www.oecd.org/sti/inno/Enhancing-Public-Research-Performance.pdf>

QAA Scotland Evidence for Enhancement Evaluation progress and planning. Resources available

online at: <https://www.enhancementthemes.ac.uk/evidence-for-enhancement/evaluation-progress-and-planning>

STEER *Your evaluative mindset*. Resources available online at:

<https://blogs.shu.ac.uk/steer/evaluation/your-evaluative-mindset/>

TASO Evidence and evaluation. Resources available online at: <https://taso.org.uk/evidence/>