Foundations of Evaluation

Resource Pack for SRA Online Course

Prof David Parsons, FRGS, FHEI, FAcSS
(Leeds Beckett University)

INTRODUCTION

This resource pack provides take away resources which provide more detail and illustrations of some of the key content of the course. There are five resources:

- *Resource A: The ROTUR framework guide for managing evaluation expectations*
- *Resource B: Jargon Buster*
- *Resource C: A summary of types of evaluation*
- *Resource D: Case study of an information matrix*
- *Resource E: The 8-step gap analysis approach to determining and sourcing evaluation information needs.*

Each resource supplements the content of the course.

We hope these resources will be useful when participants come to apply the knowledge and skills developed through the course in their own practice. Participants are welcome to share these resources with colleagues but any use should acknowledge the authorship and copyright of the resources.

Resource A: The ROTUR guide to managing evaluation expectations

A framework approach to avoiding avoidable pitfalls … some do's and do not's for reviewing Role(s) – Outcomes – Timing – Use/users – Resources.

1. Roles and responsibilities (pre-evaluation)	
DO …	**NO NOT …**
- Start at the end; who is the end-user (any intermediaries); how/when are they to be engaged in decision making	- Forget to identify internal/external procurement needs (may affect sign off; funding limits; close-open tender; marketing; etc)
- Establish who has delegated responsibility for specification (incl. objective setting; timetable; resourcing and budget)	- Delay review of info./data access needs (may affect timing; likely to need negotiating or disclosure agreements pre- start up)
- Agree who manages all aspects of sign-off/commissioning and (if different) who project manages (incl. external contractors)	- Any internal roles (including project management) will need prioritised resourcing for evaluation to deliver on-time
- Agree focus of/how much method guidance to give to contractors pre-commissioning (and who answers queries)	- Forget credible findings may need independent analysis or validation (may affect resourcing and timing)
- Establish needs for any formal steering or progress review (what for, when and who)	- Forget to brief those recruited to steering on goals and agenda; their roles and any 'rules of engagement'
- If internal evaluation: Identify who fills what roles for direction; design; delivery/data collection; analysis/verification; reporting	- Ignore the need for an evaluation champion – who will have the role of advocating change against findings (and with who).

2. Outcomes needed of the evaluation	
DO …	**NO NOT …**
- Critically review your overarching aim statement for the evaluation – is it clear, easily unambiguous and credible	- Defer seeking wider agreement on the aims and objectives; aims must precede decisions on design and are not retrofitted
- Critically review the subsidiary objectives – are they consistent with the rationale for what is being	- Forget to use objectives to clarify/set out the specific areas where evaluation evidence is needed to aid decision making

evaluated; is it coherent with any logic chain/theory of change for the 'intervention'	
- Critically review the coverage of the objectives – do they un-necessarily overlap or duplicate each other; if so consolidate	- Use objectives (what/how evidence is to be used) to set out 'method' goals (ie how to get evidence); method guidance (or prescription) follows objective setting
- Use critical review to establish any gaps in aims/objectives; is anything missing. How do aims etc change to reflect any gaps	- Hold back from asking for clarification or challenge – setting solid and appropriate expectations are the foundation of effective and usable evaluation
- Assess realism of aims and objectives; goals of the evaluation need to reflect the context, time and resources available	- Extend the aspiration for the evaluation beyond the needs of the aims and objectives; information and evidence is a tool and not just 'nice to know'
- Assess viability of aims and objectives; are they consistent with likely information availability or evidence which can be gathered*****	

*** An aide memoire on information availability**:

a) What evidence is (or is likely) to be available	*Is past documentation being used (is it subject to confidentiality constraints)? What available data/useful evidence is readily available (when/lag times to collate)? Can it be harnessed (eg are classifications/time series/updating suitable? Is there baseline/comparative evidence; etc.*
b) What 'accessible' sources (internal or external) can be used	*Past evaluation/reviews/reports; contract compliance, funding or expenditure reports; beneficiary or participation records; in-programme Mi; practice case studies; etc.*
c) Are they viable?	*Data Protection issues (identify/personal info.) may hold back accessibility/use; anonymity; is data already collated; is it machine readable; etc*
d) What are the gaps; residual information needs (from a – c)	*Best focus for 'primary' evidence collection to update/extend/add to the available evidence set against aims and objectives of the evaluation?*

3. Timing and delivery	
DO …	**NO NOT …**
- Take account of 'upstream' needs (eg internal and/or external sign-off of specification); procurement notice period (eg OJEU); marketing/tendering/commissioning decision-making lags; etc	- Forget to allow enough time also for potential contractors to produce viable tenders (2-4/5 weeks depending on needs)
- Build in 'engagement time' to liaise with stakeholders (ie specification/pre-start-up; during evaluation/steering; pre-reporting incl. findings previews; review and sign-off of reports)	- Assume stakeholders are best held at 'arms-length' until findings; earlier engagement brings challenges/delays but can help later with credibility of findings

- Allow appropriate time for sensible measurement of outcomes (and impacts) – these may take time to be realised; compressed timeframes may miss/under-represent achievements	- Skimp of time for design, testing and clearance of evaluation 'tools'; rushed design compromises information quality and reliability
- Allow sufficient time for gathering any new/additional evidence (eg survey response/reminder time) and thorough analysis and interpretation by evaluators	- Forget 'good' evaluators will need time for verification of the evidence they do collect; verification also adds to quality and credibility
- Build in time for staged/mid-point review (eg via contract review or steering); this is especially important for formative evaluations	- Under-estimate the amount of time needed for staged review within 'formative' evaluations (especially where steering groups are involved)
- Allow for 'downstream' time after (draft) reporting to review, reflect on (consult?) and sign-off evaluation before getting results/implications to decision makers etc.	- Under-estimate time needed downstream to build credibility and confidence (and understanding) of findings among intermediaries, stakeholders/doubters); evaluation utility may depend on this

4. Use and users of the evaluation	
DO …	**NO NOT …**
- Focus the evaluation approach, scope, timing and communication on the *primary* user(s). This will have been agreed from 'roles and responsibilities'; BUT …	- Forget the *secondary* users … appropriate engagement will help build credibility and also utility of the findings; are there other (non-user) stakeholders who also need to be engaged
- Clarify pre-specification how the evaluation findings are to be used; are there any expectations of change/improvement of the intervention during or after implementation, etc	- Forget that different users (primary and secondary) may have different expectations of the evaluation and its utility; unrealistic expectations of change need to countered/conditioned for all
- Identify critical timings/decision making points and align scope and approach to meet these (where appropriate)	- Forget that compressing the approach/scope to meet decision making schedules may mean compromises need to be agreed with evaluation aims/objectives' re-engineer as appropriate
- Identify if there are critical 'user' intermediaries (people, functions or bodies between whoever is accountable for the evaluation (and reporting its findings) and decision-makers	- Under-estimate the importance of champions/brokers of the evaluation findings (positive and negative) in influencing change; findings rarely speak for themselves among decision-makers
- Identify sufficiently early if/what communication strategy is needed to bring findings/implications to the user-chain.	

5. Resourcing the evaluation appropriately	
DO …	**NO NOT …**
- Recognise that resources are your budget, staff and time; these will vary with needs for internal or external evaluation	- Under-estimate the staff resource and range of skills needed for internal evaluation; external advice or peer review may help build your confidence where the skills mix/experience is limited
- Remember that 'appropriate' resourcing is led by scope, needs and expectations of evaluation – not availability of budget/time etc. Limited resources may need compromises to aims etc	- Be funding-led (what can we do for the money); critically review if the budget available is appropriate for the aims and objectives (and/or proposed approach/scope)
- Appropriately resource project/contract management; this takes time to do well does the allocated staff member have the necessary availability, skills and experience	- Forget that project managers will need to balance the added demands of evaluation management with their other tasks/roles; does the new role have clear prioritisation/sign off
- Are internal or partner interests/functions 'bought in' to resourcing decisions (eg is procurement able to support the necessary timetable)	
- Set up appropriate review/steering arrangements pre-evaluation with clear briefing on roles/responsibilities to ensure engagement and continuity across evaluation	
- Ensure timing challenges are reflected in agreed timetable (see all in 'Timing and delivery' (3) above	

Foundations of Evaluation

Resource B: Jargon Buster

Additionality: The planned outputs or outcomes occurring from an intervention which are over and above what was expected

Analytical reliability: A common test for reliability in evaluation evidence which is concerned with demonstrating any significant (in) consistencies in data preparation, processing and/or validity testing affecting confidence in the evaluations analysis and findings.

Attribution: An analysis within impact evaluation which measures or estimates the extent to which the intervention being assessed was responsible for the outcomes and impacts being measured.

Before and after analysis: A simple method of estimating the 'counterfactual' (see below) which contrasts outcomes for selected impact indicators during or after an intervention with parallel data on the same indicators before the intervention started.

Causal analysis: An analysis which isolates that part of an observed impact from an intervention which can be directly attributed to the implementation (set against other influences on change).

Comparative group (and analysis): A 'quasi-experimental' method of impact evaluation which assesses causality by contrasting specific outcomes or impacts in an intervention group with a closely matched *comparison* group such as a like-for-like geographical area.

Control group (and analysis): A method of impact evaluation which assesses causality of specific outcomes or impacts related to an intervention by setting up a 'non-intervention' group which is precisely matched (to the intervention group) and randomly selected to avoid any selection bias risks; typically in a Randomised Control Trial.

Counterfactual analysis: An analysis within an evaluation design which identifies what would have occurred (egg to outcomes or impacts) if an intervention or activity had not been implemented; comparing this to the measured outcomes after the intervention. This alternate reality is called the 'counterfactual'.

Deadweight: An effect of an intervention where (some of) the activity or benefits of an intervention would still have occurred if the intervention had not gone ahead. This is usually regarded as an unintended consequences and can be assessed for 'activity' deadweight (where actions or activities provided in an intervention would have happened without the intervention having occurred) and 'impact' deadweight (where some of the consequential changes resulting from an intervention would have been achieved irrespective of the intervention taking place. Understanding deadweight is important to impact and economic evaluation because it helps to understand value for money of an intervention.

Displacement: An unintended consequence of an evaluation where the positive outputs or outcomes of an evaluation are offset by negative outputs or outcomes occurring elsewhere (eg participants in a new course or programme recruited from those planning on entering an established course).

External validity: A demonstration that the methods used and the evaluation results coming from these can be confidently 'generalised' to another similar (intervention) context or situation. Evaluations with strong external validity are said to provide 'transferable' evidence.

Hybrid evaluation: An evaluation methodology using a mixture or evidence collection methods (mixed mode) and typically combining quantitative and qualitative methods to tri-angluate (see below) different evidence perspectives or sources.

Impact: An observed effect resulting from an (evaluated) intervention and as a consequence of delivering or achieving specific activities or 'outputs'; usually associated with measurement of longer term 'consequential changes' from the intervention.

Internal validity: The focus for demonstrating validity of evaluation evidence as a trustworthy reflection of what is being evaluated. This will usually set out measures of statistical confidence (of 'new' quantitative data) and an assessment of (any) bias, distortions or variability in evaluation evidence which affects confidence in the findings.

Knock-on effects: An unexpected, unintended or indirect consequential effect of an (evaluated) intervention.

Leakage: Effects within measured outcomes or impacts which support others outside the targeted or expected intervention group (eg eligible age group of participants or geographical area of intervention).

Measurement reliability: A common test for reliability in evaluation evidence which is concerned with acknowledging any significant (in)consistencies in the use of (different) processes, indicators and tools used to gather and measure information for an evaluation.

Observer reliability: A common test for reliability in evaluation evidence which is concerned with any significant (in) consistencies that may be due to using different interviewers or observers (raters) when collecting evidence. This is also called 'rater' reliability.

Outcomes: An early or short term 'impact' resulting from an (evaluated) intervention and usually resulting as a consequence of delivering or achieving specific activities or 'outputs'.

Participatory evaluation: An approach to evaluation conduct based on, but narrower than, Participatory Action Research (PAR) principles which provides opportunities for evaluators to put stakeholders, including beneficiaries, centre-stage in evidence-collection and review.

Primary evidence: Quantitative and/or qualitative evidence in an evaluation which is generated directly by the evaluator (or on their behalf) from additional information collection methods such as participant or practitioner interviews or surveys.

Process tracing: A qualitative technique developed within 'generative causation' for developing and using in-depth user/impact case studies to 'trace' the evolution of impacts as the engagement in the implementation evolves.

Proportionality: The principle of evaluation design which sets out that in addition to the need for reliable information, the choice and mixture of evidence gathering and analytical methods used should be 'proportionate' to the objectives, scale and nature of what is being evaluated.

Secondary evidence: Quantitative and/or qualitative evidence in an evaluation which is collated from existing sources of evidence within or outside an intervention including from, for example, management or monitoring information, past research (or evaluations) and available documentary sources.

Situational reliability: A common test for reliability in evaluation evidence which is concerned with acknowledging any significant (in)consistencies in the conditions or circumstances in which evidence is gathered and which might cause variations in data quality.

Small 'n' evaluation: A small scale evaluation where 'n' relates to the overall size and scope of participation in what is being evaluated perhaps in a trial scheme, small-scale pilot, highly localised or single site intervention or one involving a narrow or specialised beneficiary group.

Social Return on Investment: SROI is a specialised method developed first in the area of social enterprise and building on cost-benefit analysis aimed at valuing social and environmental impacts from initiatives and actions and which may not be fully covered in more conventional approaches to economic evaluation.

Spill-over effects: Unplanned consequences arising from (evaluated) interventions and activities and which can be positive (adding to the quality and range of expected impacts) or negative (detracting from programme achievements and impacts).

Subject reliability: A common test for reliability in evaluation evidence which is concerned with identifying any significant (in)consistencies due to a contrasting focus or different quality of data gathering in evaluation subjects.

Substitution: Measured outcomes or impacts (or aspects of them) on an intervention group which are realised at the expense of others outside the intervention group, often as unintended consequences from the intervention (eg. New employment support actions for a disadvantaged group resulting in existing support actions being closed or run down).

Tri-angulated evidence: A commonly used evaluation approach providing validation of both quantitative and qualitative evidence through cross verification from two or more sources, typically derived from combination of several research methods in assessing the same phenomenon.

Unintended consequences: Unexpected impacts and effects of (evaluated) interventions and activities which need to be identified and taken into account in any assessment of net impacts. See also spill-over effects.

Valuation: Techniques for measuring or estimating the monetary and/or non-monetary value of observed outcomes and impacts, contributing to understanding added value or cost-effectiveness of the evaluated intervention.

Resource C: A summary of types of evaluation

Type of evaluation	What is it?	Where is it useful?
Process evaluation *Also known as:* • *Managerial evaluations,* • *Developmental Evaluations,* • Implementation evaluation • *Programme Evaluation.*	These focus on evaluating the pathways and mechanisms through which an intervention takes place. This is usually with a view to seeing how they could be improved, rolled out or transferred to similar situations. Process evaluations can be called various things but whatever the label used, they share an emphasis on giving policy or other decision-makers evidence of how (well) an intervention has been implemented or managed, how it operates, and how it produces what it does, rather than an understanding of the changes it has (or has not) brought about or its wider effects. They typically use a range of quantitative and qualitative methods. They will often cover subjective issues (such as perceptions of how well a policy has operated) and objective issues (the factual details of how an intervention has operated, typically using administrative data, where available).	Typical uses of process evaluations include: • Accountability review (for expected outputs) • Review of efficiency of delivery for intervention inputs, activities and/or outputs • Examining programme (contractor) delivery and achievement of activity or output targets • Assessing transferable potential for pilot or trial programmes • Assessing lessons from what worked well and less well, why and what could be improved? NOT suitable for: • Evaluations looking at what outcomes or impacts (consequential changes) have been achieved • 'Causal' assessment of intervention outcomes

Impact evaluations *Also known as:* • *Randomised Control Trials,* • *Comparator Evaluations,* • *Experimental Evaluation* • *Quasi-experimental Evaluations,* • *Theory-based Evaluations.*	The central question of any evaluation is 'what difference has the intervention made'? This type of evaluation is concerned mostly not with the activity or outputs of an intervention (as a process evaluation would be) but with measuring or estimating the *consequential change* that results from those activities/outputs. This usually has a quantitative focus but decision makers are often concerned with rather more subjective issues such as 'why', what were the success factors and setting what the results were against the expectations of what needed to change to address a problem or challenge. A wide range of approaches with different pro's and con's can be used to go about measuring 'consequential change'. Picking the right one is often a challenge, and especially where interventions take place in complex situations where their outcomes and impacts may have also been contributed to by influences from outside the intervention itself. Impact evaluation is consequently also concerned with looking at the *attribution* of impacts to what was done by the intervention itself; this uses a range of approaches to measuring the 'counterfactual' – or what would have happened if the intervention had not taken place.	Impact evaluations can use different sorts of so called 'experimental' approaches to measure or estimate outcomes and causality, or can take an approach focussed more on unpicking its 'results' set against its underpinning assumptions and expectations – theory based methods. Either can work well in the 'right' circumstances and will be aimed at answering questions about, for example: • What measurable short or medium term outcomes or longer term impacts occurred (both direct and indirect)? • Where there any unintended consequences which need to be taken into account? • How much of these outcomes or impacts can be attributed to the intervention? • Have different groups been impacted in different ways, and why? • What aspects of the intervention enabled (or constrained) outcomes – and success? • What aspects of the intervention context influenced outcomes?
Economic evaluations *Also known as:* • *Value for money evaluations* • *Cost-benefit evaluations*	A tightening public purse and pressure on margins and the 'bottom-line' in many organisations mean that evaluations may need to centre on if the intervention was worth its investment (funding); did it provide value for money (VfM)? These will involve different approaches to economic evaluation to review if the costs of an intervention are	Economic evaluations can be used where decision makers need to go beyond looking at impact to understand if intervention outcomes are justified and whether the intervention remains the most effective use of resources. Approaches to economic evaluations come in different shapes and sizes and may variously look at:

	outweighed by its (monetised) results and the benefits it produces. While 'VfM' itself involves some subjective assessments, economic evaluation is essentially quantitative and focussed on measuring the costs of resource inputs, and usually setting these against outputs or outcomes to assess an intervention's added value. Often this will need some comparative element to provide information about cost-effectiveness, usually in money terms. Economic evaluations are not cost-benefit assessments but often draw on similar measures (eg Value for Money Indexes) and methods of *monetising* inputs, outputs and outcomes.	• Investment (cost) description and distribution • Cost-effectiveness of the intervention • Cost-utility – setting the investment against a 'usefulness' benchmark eg (value of a new job created to the economy) • Cost-benefit evaluation – involving a series of cost-benefit assessments based on different assumptions about monetised outcomes. A full cost-benefit evaluation is a complex undertaking and would usually go beyond looking at monetised outcomes to also compare the benefits and costs of other ways of achieving the same results.
Meta-evaluation	This is an evaluation of, or across, other evaluations, using evidence which has already been conducted. A meta evaluation is more than a literature review since they are focussed on past evidence-based assessments and not commentary or opinion-based material. Meta evaluations involve a formalised and structured approach to looking across, and comparing, past evidence often using full or condensed systematic review type methods (eg rapid evidence review) drawn from previous evaluations. The comparisons will also need to look not just at what that past evidence was saying but how it was gathered, its validity and reliability.	Meta evaluation approaches can be stand-alone evaluation activities or may be used as prelude to a subsequent experimental or theory-based evaluation of a particular intervention. They are useful where programme planners or policy makers want to learn from past experience and where 'cumulation' of past evidence can help to better understand or explain wider processes. NOT suitable where • There is insufficient/inadequate (eg poor quality) past comparable evidence; Or • Where available past evidence has insufficient comparability in its timing, focus or scope to produce useful meta-analysis.

Resource D: Case study of developing an *information matrix*

Introduction

The case study is of an 'ex ante-evaluation' conducted of problem gambling (PG) in Leeds City region ahead of the development of the UKs 4th 'super-casino' as part of the redevelopment of the city's Victoria Gate area. The evaluation followed a city region wide consultation which showed community concerns over rising levels of problem gambling, and aimed to inform decisions about future impact evaluation and for (any) improved provision of harm minimisation support to problem (and at risk) gamblers. Specific objectives were to:

1. Establish a socio-demographic baseline for better understanding comparative gambling prevalence in the Leeds city region
2. To review the supply and demand for gaming activity in Leeds focussing on 'land-based' provision including comparative prevalence
3. Review the comparative scale and distinctiveness of problem gambling in the Leeds city region through an appropriate benchmark, and the known harms and adverse impacts on problem gamblers
4. Establish the access and supply of preventative & remedial support to problem gamblers in Leeds City region including known gaps or constraints to provision and take-up
5. Provide recommendations for a future evaluation of the (super casino) development to assess its impact on gambling prevalence, problem gambling and implications for any enhanced supply or access to harm minimisation support services.

The evaluation

The evaluation was to be conducted independently and intensively (26 weeks) and reported to a cross-stakeholder steering group chaired by Leeds City Council. It was commissioned from a multi-disciplinary team at the Carnegie Faculty of Leeds Beckett University and used a quasi-experimental design to support the comparative aspects of the evaluation.

Information matrix

An early development from the evaluation team was an information matrix produced as a working draft and subsequently as a (revised) evaluation template (as below) by:

- Subdividing a number of the evaluation objectives into separate evidence gathering contents (eg supply vs demand of gambling) to create 8 distinct 'objectives' groups (1, 2 a and b), 3a and b, 4a and b, and 5)
- Establishing 'core information needs' for each of those groups – with 30 core-info sub-sets
- Mapping the aggregate subsets against likely evidence sources (work packages – WP1-5) – some overlapping.

Following discussions at the steering group and (identify some additional potential evidence sources) this guided the development of the full evaluation methodology, its articulation and timing.

Information Matrix

Information needs (Work Package for evaluation)	**WP1**	**WP 2.1**	**WP 2.2**	**WP3**	**WP 4**	**WP5**
Evidence focus/method	*Secondary analysis of national data-sets; local data*	*S'holder & sector (op's) semi-structured interviews*	*PG semi-structured interviews*	*Rapid Evidence Review of vulnerability factors*	*Qualitative evidence of Impacts from s'holders and PGs*	*Survey of PG support services + follow-up agency interviews*
Evidence coverage/scope	*Comparative PG no's and demographics: Leeds, city region; 6 QE comparators*	*PG profiling experience/suppl. data (evidence & experiences)*	*PG profiling: lived-context, needs/utility of support*	*Research evidence and benchmarking analysis from secondary sources*	*Interview identification of wider PG impacts*	*Mapping LCR PG service supply/referral, demand data, 'quality & gaps*
1. Socio-demographic baseline: a. Spatial distributions and pop density b. Socio-demographic (age profile, ethnicity, etc characteristics) c. Socio-economic profiles (incl. unemployment; u/e durations) d. Etc.	√					
2a. Gaming supply in Leeds: a. Mapping operations, operators and activities; b. Licensing (overall, venues – current/trends)	√	√				

c. Licensing by activity (casino, bingo, LBO, AGC/FEC) d. Cross-activity relationships e. Trends and developments; etc						
2b. Gaming demand & prevalence in Leeds+ and comparative: a. Gambling/player distribution, volume and participation b. Preferences and activity demographics; c. Frequency, reasons, etc d. Risk factors and behaviours e. Locational evidence/clusters – spatial profiling; f. Spending.	√	√				√
3a. Problem gambling benchmark/Leeds distinctiveness: a. Scale and PG characteristics (DSM IV AND PGSI); b. Known distribution and activity; c. Vulnerability & risk behaviour; d. PG determinants.	√	√		√		√
3b. PG impacts and effects: a. Identified social impacts (debt., family/relationship effects/breakdown, etc); b. Offending and anti-social behaviour;	?	√			√	√

c. Economic activity including employment; d. Health and welfare; e. Other non-specified impacts						
4a. Supply of preventative & remedial support in Leeds: a. Education and awareness (?); b. Advice and counselling (including indirect services); PG/addiction treatment; c. Informal support (eg networks); etc.		√				√
4b. Access to preventative and remedial support: a. Direct access; b. Gaps/access constraints; c. Cross-agency referrals and inter-relationships; etc.		√				√
5. Recommendations: a) For evidence needs for future impact evaluation of super-casino development b) For evaluation approach, focus and timing c) For addressing any gambling harm minimisation access or support gaps or remedial provision in the city region.	√	√	√	√	√	√

Source: Kenyon A, Omerod N, Parsons D J, Wardle H (2016), *Problem Gambling in Leeds: A Research-based Evaluation for Leeds City Council and Partners*. Leeds Beckett University.

Foundations of Evaluation

Resource E: The 8-step gap analysis approach to determining and sourcing evaluation information needs

Why the 8 step approach?

Whatever it's 'type' it is unlikely (although certainly not impossible) that all the information needs of an evaluation will be delivered wholly by 'secondary' sources of information, and analysis. 'Primary' sources are used to fill the gaps between what is evidence is available (and appropriate) such as from programme monitoring data, and what is needed to provide a sufficient evidence-base to address the evaluation objectives.

For the evaluator, the options for using primary evidence-gathering sources will usually greatly exceed the potential to exploit them within a specific evaluation. Beyond a possible plethora of choices is also the reality that primary data will need specific and usually customised evidence gathering methods, so is likely to be both costly and time consuming. Consequently, for most evaluations hard choices need to be made about what primary evidence is needed and can be viably (and validly) collected.

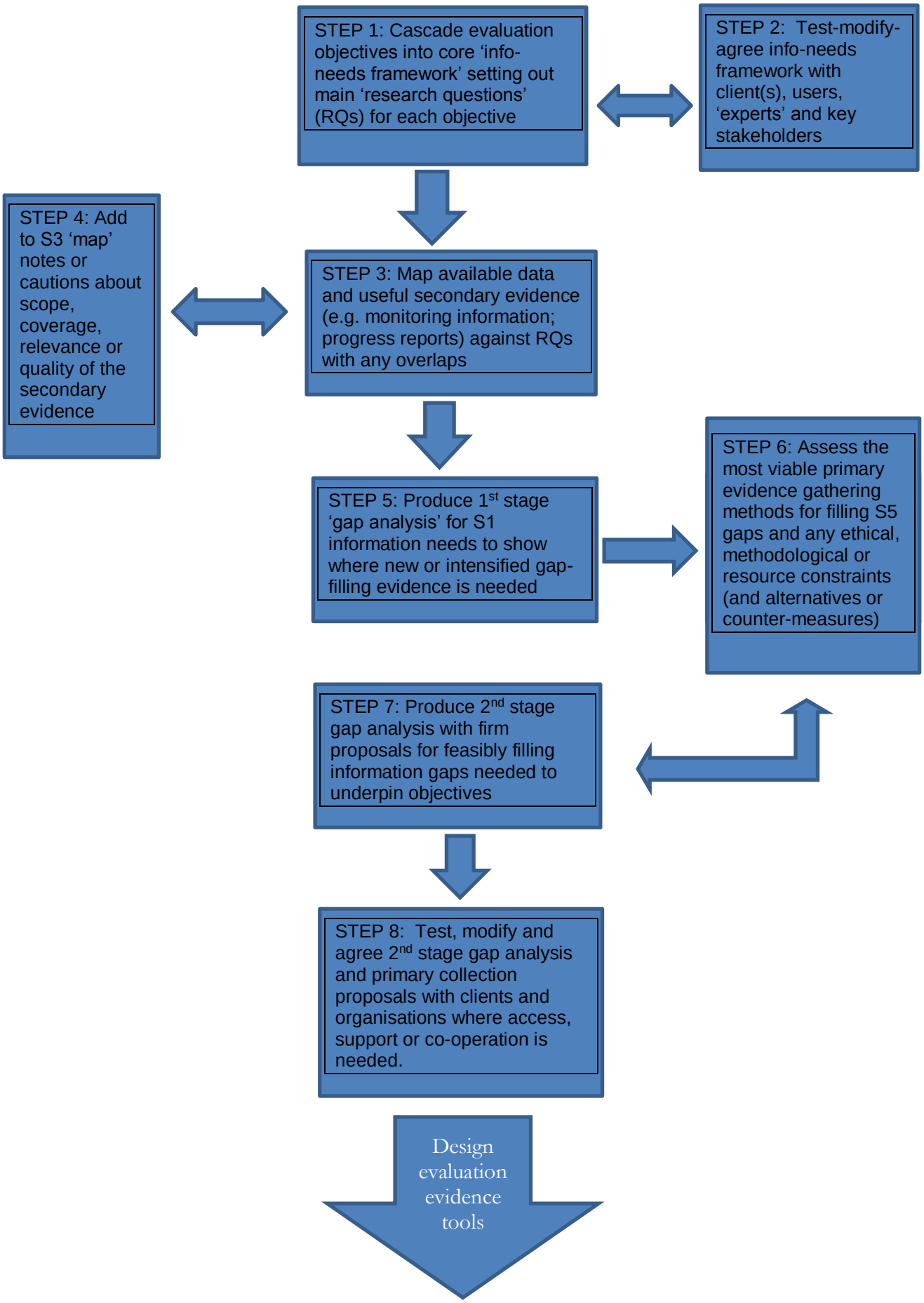
The expectations of the evaluation (and often the specification itself) may already have made some assumptions about this. Calls in the evaluation objectives, for example, for participant surveys or interviews, stakeholder review, or 'best practice' case studies, may have as much to do with aspirations than sound methodological judgement, but they provide a starting point. A more systematic basis for assessing where primary evidence is needed is to use a "gap analysis" approach, perhaps based on the 8 stage model set out here. There are (many) other ways of tackling this but this one is tried and tested and will be found to fit all evaluation types and most circumstances.

How to apply the 8-step model?

The 8 step model is summarised in Figure 1. It is a practical and progressive approach to defining primary evidence needs which is founded on the context, aspirations and needs of the evaluation and the requirements of users. This involves working through the eight stages to identifying evidence gaps and from this shaping the priorities and methods needed to fill gaps within resources. It is consistent with collaborative approaches to evaluation and builds in review and refinement stages which involve users and perhaps others stakeholders providing staged feedback as it develops.

Step 1: Like any progressive approach to methodological development its effectiveness depends very largely on its foundations. This involves starting with clear and realistic objectives for the evaluation which are then cascaded into core information needs, sometimes referred to collectively as the 'research questions' (RQs). Each of the evaluation objectives will be broken down into a number of specific RQs.

Figure 1: The 8 step model approach to a primary evidence gap analysis



It does not matter if a few of these core information needs (RQs) overlap more than one objective; what is important is that put together they set out an overarching *core information framework* covering for each of the objectives the specific themes and issues that need to be addressed to meet the overall evaluation goal.

Step 2: This is the preliminary feedback opportunity. It involves taking the information agenda from stage 1 to users and, as appropriate, other knowledgeable stakeholders, to test for its comprehensiveness. Feedback may identify modifications to the RQs, suggest other information needs and/or provide feedback on information priorities. It is often the case that this early feedback shows up constraints in the way the evaluation objectives themselves were set out; this is much better to find out early in this process than when the evidence gathering is already underway when it may be too late to adapt. This feedback will allow for refinement of the core information needs framework.

Step 3: The availability of likely secondary sources is here mapped against the (revised) core information framework to show where available evidence will meet or contribute to some of the RQs. This is not a simple mapping exercise; it involves sense checking the appropriateness of the secondary sources that might be used to meet needs against the requirements of them. Does the available evidence cover the required evaluation units (eg people, specific age or participant groups); is it of sufficient currency, scope, scale and focus; does it classify evidence in ways that are appropriate to the evaluation; will it be available in the format and timeframe needed; etc?. If not, the secondary evidence may be 'available' but will fall short of being 'appropriate' for the evaluation needs; as such it will still constitute an evidence gap.

Step 4: The evaluator may not have a monopoly of wisdom on what sources are available and appropriate. Assumptions may be being made about coverage or timeliness of, for example, available data about programme participants, activity monitoring or associated management information which are out of date or inaccurate. To counter this risk, the stage 3 map of secondary evidence against the RQ elements is subject to scrutiny by users and also any stakeholders likely to be contributing that evidence (eg from monitoring systems or existing databases) to pick up any inaccuracies. Success at this stage is about the quality and comprehensiveness of the secondary evidence mapping. It is the point where the evaluator would be best served by finding out that monitoring data (perhaps to be used for survey sampling) does not cover all participation in the programme or initiative to be evaluated, or perhaps it will not be available in the timeframe needed.

Stage 5: With the RQ-secondary evidence map updated (from stage 4 feedback), the process had produced a preliminary gap analysis of the core information needs. This shows what required evidence cannot be met adequately from existing sources and will require additional (primary) data collection to meet those needs. This 1st stage gap analysis will be supplemented by identifying likely evidence collection methods to fill the various gaps; recognising that some ways of collecting evidence (eg surveys; interviews) may use a single method to cover a number of RQs. Ideally, the 1st stage gap analysis will be best served by giving some idea of scale of the necessary methods (eg survey of a probability sample of c.1-2,000 individual beneficiaries; individual case studies of cross-section of c.30-40 individuals; etc). At this stage, depending on the information gaps apparent from stage 4, it may be obvious to the evaluator that the emerging requirements for primary evidence collection may exceed available budget or timeframe. The 1st stage gap analysis is, however, an *idealised* framework and in these situations will need to be supported by the evaluator's assessment of likely viability within the scope of the evaluation.

NB. Just as with secondary sources, the evaluator using this staged approach will need to be familiar with the range of qualitative and quantitative methods that can be employed to fill crucial gaps and be well placed to propose appropriate choices among different possibilities for capturing evidence that best fits needs.

Stage 6: This is the penultimate and crucial feedback loop with users and perhaps selected stakeholders. It will test the idealised evidence gap assessment from the 1st stage gap analysis and also the viability of the primary methods needed to fill those gaps. If that gap analysis also sets out constraints of budget, timing or other resources to the idealised method proposals it will also involve asking users to make value judgements to prioritise or streamline the information needs by assessing what is 'crucial and viable' against 'what is 'desirable'. This in turn may require some modification to the evaluation objectives (or scope) to align the information requirements against available resources.

NB. If time is short, it may be possible to compress stages 1-5 into a single sequenced set of activities conducted within the evaluation team but without feedback to users (stages 2 and 4). Here, feedback will then rely wholly on stage 6 to pick up on invalid assumptions as well as viability of primary methods proposed to fill gaps. The compressed approach will save on time but at the risk of some substantial re-engineering of primary evidence proposals if earlier assumptions prove flawed.

Stage 7: A 2nd stage gap analysis is produced following feedback at stage 6 with revised (primary) method mix for addressing those gaps. This may involve a substantial revision to the information needs and/or method mix if the 1st stage idealised method mix is subject to substantial resource constraints. It will also involve tightening up on the primary evidence method proposals to show, for example, broad approach and scale of activities, selection and access issues.

Stage 8: The concluding stage involved the final feedback from users and knowledgeable stakeholders on the 2nd stage gap analysis and method proposals. It is likely to include any partners or third parties from whom the evaluation will need to agree access to data, sites or, for example, survey participants or interviewees. Feedback will provide for the final evaluation design and for the evaluator to progress towards designing any specific evidence collection tools informed by a clear (and agreed) agenda of what they needs to cover. It also provides a platform, if needed, for agreeing information or other access protocols with partners or third parties to minimise the risks of disruption to later evidence gathering.

Who to engage in feedback (stage 2, 4, 6 and 8)?

Overlaying this approach is the quality of engagement that can be brought to bear on testing and enhancing the proposals. Testing is important not just in sharpening the gap analysis but, in due course, in providing users with enough information about what is being collected (and how) to judge the validity of subsequent conclusions and recommendations.

The principal users (and funders) of the evaluation are the most likely to need to be involved, and also those on whom the evaluation will be reliant to provide access to, for example, key data sources or databases to provide access to programme or initiative participants (and perhaps providers and practitioners). To this might usefully be added 'knowledgeable others' among stakeholders or experts who have particular experience (or expectations) to share.

Who should be engaged will not just be a technical exercise; it will be a political one. Users may have distinct (and possibly overriding) views about who should be involved (or should not).

How long does this take?

Most evaluations will be under pressure to ‘get going’, and evaluators may feel they have little time or scope to work through this staged approach. In fact, the 8-step approach need not take a lot of time; it can be compressed (as noted above) and it can also be conducted intensively where the evaluator is confident they can secure appropriately fast turnaround of feedback. It can also be run in parallel with other aspects of evaluation set up such as expectations review, objectives clarifications, inception and stakeholder briefings.

The (short) investment in time that using this staged approach takes at the start of the project is returned by reducing risks of disturbance to data collection later in the evaluation. It is also a starting point for building stakeholder understanding of what is being done and why in the evaluation and this provides a platform for increasing confidence and credibility in its eventual findings.