



Options and Choices in Impact Evaluation

Resource Pack for the SRA Course

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

Social Research Association Workshop

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 four resources:

- *Resource A: Evaluation types and best use*
- *Resource B: Using the ROTUR guide for managing expectations*
- *Resource C: Ready reckoner for experimental approaches to evaluation*
- *Resource D: Some analytical methods for alternative evaluation approaches*

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 Unpicking evaluation types and purposes

Type of evaluation	Typically for the purposes of ...	Good for ...	Not so good for ...
<p>Process evaluation:</p> <p>Evaluating the mechanisms through which a responsible gambling action or intervention takes place.</p>	<ul style="list-style-type: none"> • Providing evidence of how (well) an intervention has been implemented or managed against expectations (budget/targets?) • Reviewing how it operates, and how it produces what it does and differences in effectiveness • Identifying what works well (and does not), for who/where/when and improvement potential • Assessing cost-effectiveness and areas for cost-efficiencies. 	<ul style="list-style-type: none"> • Cost-accountability • Roll-out or scale up potential of a trial/pilot • Understanding ‘what works’ and why • Staged or formative evaluation in a longer term initiative to identify improvement potential 	<ul style="list-style-type: none"> • Looking at outcomes or impacts (see impact evaluation) • Assessing any aspects of value for money (see economic evaluation)
<p>Economic evaluation:</p> <p>Evaluating the costs of inputs, outputs or outcomes or overall value of an action.</p>	<ul style="list-style-type: none"> • Measuring costs and cost-efficiencies against business plans/budgets or other expectations • Quantifying cost-efficiencies and cost-effectiveness in money terms • Measuring or estimating value for money or value-added of a responsible gambling action (or set of actions) 	<ul style="list-style-type: none"> • Accountability (assessing costs against budgets) • Projecting cost-efficiencies or cost-utility • Reviewing cost-benefits of outputs or outcomes in money-terms 	<ul style="list-style-type: none"> • Pilot, staged or formative evaluations • Where outcomes cannot be credibly converted to ‘money’ values
<p>Impact evaluation:</p>	<ul style="list-style-type: none"> • Quantifying outcomes (short or medium-term) or impacts (longer term) resulting from a responsible gambling initiative 	<ul style="list-style-type: none"> • Measuring (or estimating) both ‘hard’ and ‘soft’ outcomes and impacts 	<ul style="list-style-type: none"> • Interventions lacking clear expectations of impact(s)

<p>Evaluating the outcomes or impacts (<i>consequential change</i>) resulting from an intervention set against its aspirations.</p>	<ul style="list-style-type: none"> • Unpicking impact contrasts within diverse groups (eg participants) • Identifying unexpected (additional) impacts or unintended consequences • Assessing the contribution made by an initiative to overall outcomes/impacts (ie causal attribution) • Understanding impact determinants and success factors (enablers) and constraints • De-constructing design and contextual influences (from theory-based approaches). 	<ul style="list-style-type: none"> • Taking account of change lag effects (quantifying outcomes over time) • Assessing how outcomes and impacts come about • Critically assessing achievements against expectations • Demonstrating effectiveness to stakeholders 	<ul style="list-style-type: none"> • Short term or very intensive evaluations • Interventions without scope/potential for quantification <p><i>NB. BUT different impact evaluation approaches have different application pro's and con's</i></p>
<p>Meta-evaluation: Evaluations which draw on evidence from past research or evaluation in parallel areas</p>	<ul style="list-style-type: none"> • Setting a start-up context for a new or modified intervention • Constructing a (past) evidence-based case for a new intervention • Contributing a multi-source benchmark (eg for an ex ante evaluation) • Providing indicators of what needs to be looked at to assess effectiveness of a new intervention 	<ul style="list-style-type: none"> • Systematic use of past evidence (where there are empirical foundations) • Contributions for design of novel interventions • Situations where there has been extensive past work (to select which are most relevant). 	<ul style="list-style-type: none"> • Intensive evaluations (worthwhile meta evaluation takes time) • Situations where there is little/no parallel past experience • Where past evidence capture methods have limited external validity (generalisability) • Where past evidence is weakly constructed or mainly small-scale.

Resource B: Using the ROTUR Framework for Managing Expectations

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 evaluated; is it coherent with any logic chain/theory of change for the 'intervention'	- Forget to use objectives to clarify/set out the specific areas where evaluation evidence is needed to aid decision making
- 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'

<p>- Assess viability of aims and objectives; are they consistent with likely information availability or evidence which can be gathered*</p>	
---	--

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

<p><i>a) What evidence is (or is likely) to be available</i></p>	<p><i>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.</i></p>
<p><i>b) What 'accessible' sources (internal or external) can be used</i></p>	<p><i>Past evaluation/reviews/reports; contract compliance, funding or expenditure reports; beneficiary or participation records; in-programme Mi; practice case studies; etc.</i></p>
<p><i>c) Are they viable?</i></p>	<p><i>Data Protection issues (identify/personal info.) may hold back accessibility/use; anonymity; is data already collated; is it machine readable; etc</i></p>
<p><i>d) What are the gaps; residual information needs (from a – c)</i></p>	<p><i>Best focus for 'primary' evidence collection to update/extend/add to the available evidence set against aims and objectives of the evaluation?</i></p>

3. Timing and delivery

DO ...	NO NOT ...
<ul style="list-style-type: none">- 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	<ul style="list-style-type: none">- Forget to allow enough time also for potential contractors to produce viable tenders (2-4/5 weeks depending on needs)
<ul style="list-style-type: none">- 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)	<ul style="list-style-type: none">- Assume stakeholders are best held at 'arms-length' until findings; earlier engagement brings challenges/delays but can help later with credibility of findings
<ul style="list-style-type: none">- Allow appropriate time for sensible measurement of outcomes (and impacts) – these may take time to be realised; compressed timeframes may miss/under-represent achievements	<ul style="list-style-type: none">- Skimp of time for design, testing and clearance of evaluation 'tools'; rushed design compromises information quality and reliability
<ul style="list-style-type: none">- Allow sufficient time for gathering any new/additional evidence (eg survey response/reminder time) and thorough analysis and interpretation by evaluators	<ul style="list-style-type: none">- Forget 'good' evaluators will need time for verification of the evidence they do collect; verification also adds to quality and credibility
<ul style="list-style-type: none">- Build in time for staged/mid-point review (eg via contract review or steering); this is especially important for formative evaluations	<ul style="list-style-type: none">- Under-estimate the amount of time needed for staged review within 'formative' evaluations (especially where steering groups are involved)
<ul style="list-style-type: none">- Allow for 'downstream' time after (draft) reporting to review, reflect on (consult?) and sign-off evaluation before getting results/implications to decision makers etc.	<ul style="list-style-type: none">- 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 ...
<p>- Focus the evaluation approach, scope, timing and communication on the primary user(s). This will have been agreed from 'roles and responsibilities'; BUT ...</p>	<p>- 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</p>
<p>- Clarify pre-specification how the evaluation findings are to be used; are there any expectations of change/improvement etc</p>	<p>- Forget that different users (primary and secondary) may have different expectations of the evaluation and its utility; unrealistic expectations of change need to be countered/conditioned for all</p>
<p>- Identify critical timings/decision making points and align scope and approach to meet these (where appropriate)</p>	<p>- 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</p>
<p>- 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</p>	<p>- 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</p>
<p>- Identify sufficiently early if/what communication strategy is needed to bring findings/implications to the user-chain</p>	

5. Resourcing the evaluation appropriately

DO ...	NO NOT ...
<ul style="list-style-type: none"> - Recognise that resources are your budget, staff and time; these will vary with needs for internal or external evaluation 	<ul style="list-style-type: none"> - 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
<ul style="list-style-type: none"> - 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 	<ul style="list-style-type: none"> - 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)
<ul style="list-style-type: none"> - Appropriately resource project/contract management; this takes time to do well does the allocated staff member have the necessary availability, skills and experience 	<ul style="list-style-type: none"> - 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
<ul style="list-style-type: none"> - Are internal or partner interests/functions 'bought in' to resourcing decisions (eg is procurement able to support the necessary timetable) 	
<ul style="list-style-type: none"> - Set up appropriate review/steering arrangements pre-evaluation with clear briefing on roles/responsibilities to ensure engagement and continuity across evaluation 	
<ul style="list-style-type: none"> - Ensure timing challenges are reflected in agreed timetable (see all in 'Timing and delivery' (3) above 	

Resource C: Ready reckoner on uses of experimental, non-experimental and theory-based approaches to impact evaluation

More viable when:		Less viable when:
Experimental (RCT/QE)	<ul style="list-style-type: none"> • Allocated and highly skilled staff are available (<i>internal evaluation</i>) or substantial allocated budget for (<i>external evaluation</i>) procurement of appropriately skilled contractors for design • Evaluation focuses on a wholly new policy/initiative or intervention • Evaluation also focuses on a distinct and identifiable change in behaviour, practice or situation affecting participants • The intervention is of a clearly defined scale; best fitted to trial or pilot scale actions • The intervention takes involves only one or closely inter-linked inputs or activities and takes place in a non-complex (RCT) environment • The intervention takes place in a predictable and stable delivery environment • The intervention activity (RCT) or activities (QE) are isolatable with likely distinctive sustainable effect on predictable and measurable outcome (relative to other changes taking place) • Control (RCT) or comparative (QE) data can be designed into the evaluation with quality data for objective selections (and randomised (for RCTs). 	<ul style="list-style-type: none"> • Limited (or no) allocated staff resource or an appropriate budget available • Procurement constraints (including lead into contracting for short time frames) affect financing (or contracting) of sub-contracted expertise • Intervention to be evaluated involves an established or modified policy or initiative, one which is consolidating current activity or involves multiple effect policy goals or expected outcomes • Intervention has already started (ie where evaluation is commenced during the intervention activities) • Evaluation (or users) require a discontinuous or formative approach to assessment and reporting • Intervention context or environment is unstable or not controllable for comparative measurement over time • Intervention involves a complex environment with multiple confounding factors which may affect outcomes (or its accurate measurement) • Intervention effects or outcomes is/are likely to involve significant unknown 'external' influences • Expected outcomes/impacts are expected to take some time to be measurable due to long effect lead times or incremental impacts • The available participation data does not provide for robust and ethically compliant selection of appropriate control protocols (RCT) or valid comparators (QE)

	<p><i>NB. RCTs and QE can be used where there is a small expected effect but this will require large and costly samples for robust evidence.</i></p>	<ul style="list-style-type: none"> • The available participation data are partial, inconsistent or subject to unresolvable GDPR access issues.
<p>Non experimental</p>	<ul style="list-style-type: none"> • No or very limited provision has been made for (<i>external</i>) evaluation or where low/no cost are critical considerations (<i>internal evaluation</i>) • Intervention is already under-way or is discontinuous • Multiple influences on the intervention effects or outcomes are likely to be involved • Other designs are constrained by ethical considerations in selecting 'comparator data (NE uses 'off the peg' comparator data and consequently does not generate it directly) • Consistent 'participant' start ('before') and end ('after') or other comparator data (time/area/context etc) are readily available • Impact evaluation reporting requirement is progressive or formative as well as summative evaluation • Outcome and causality measurement methods can be used in combination (eg before and after + trajectory analysis) • Effects or outcomes are expected to be relatively large (constrained designs have lower levels of internal validity which would invalidate reliable measurement of small effects). 	<ul style="list-style-type: none"> • Users are looking for statistically demonstrable valid evidence of intervention contributions or causality outcomes (estimation of causality is less reliable than measurement) • Evidence (data) on non-intervention influences cannot be isolated to adjust analysis for unintended consequences • There is not suitable well-matched and timely comparative data • Limited analytical confidence where 'net' impact differences are small.
<p>Theory-based approaches</p>	<ul style="list-style-type: none"> • Highly skilled staff with TBE experience are available (<i>internal evaluation</i>) or substantial allocated budget for (<i>external evaluation</i>) experienced TBE contractors • User organisations have a culture of using theory of change or complex logic modelling to help set up new policies or interventions • Intervention is a novel policy/initiative or intervention and with an uncertain or unpredictable cause-effect relationship 	<ul style="list-style-type: none"> • Users are focussed on quantification of effects or outcomes and where evaluations which do not emphasise measurement will not be trusted • Users where established evaluation processes have emphasised fully (or quasi) experimental approaches (randomistas)

	<ul style="list-style-type: none">• Impact evaluation in complex 'systems' where there are multiple influences on intervention effects and outcomes (ie <i>generational</i> causality)• Interventions where there are expected to be unpredictable or volatile 'internal' influences on outcomes• Interventions where many of the 'external' influences are 'unknowns' or where their interactions with intervention processes are uncertain• Interventions where it is important to unpick the quality of the innovation (eg pilot or trial or novel actions) by testing soundness of intervention logic or expected cause-effect	<ul style="list-style-type: none">• Interventions with simple cause-effect relationships or with an established or well proven logic and mechanisms (ie little scope for TBE to add value).
--	---	---

Resource D: Some analytical methods for ‘alternative evaluation approaches’

What is it about; where does it fit?	Some pro’s	Some Con’s
Bayesian Updating		
<p>Uses Bayes theorem) to help show the extent to which the TBE evidence supports contribution claims by assessing the probability of a contribution being valid. It can use different methods to estimate simple probabilities including empirical evidence, modelling or ‘subjective probabilities’ (via consensus).</p> <p>Bayesian Updating has been an established analytical feature for other professions and can be well-placed for assessing causality of outcomes from (mainly) qualitative contribution claims. It can help put rigor back into evaluations which lack quantification options, and is suited to evaluation of multi-activity programmes in complex settings. DFID and others have used it in TBE contexts in the UK.</p>	<ul style="list-style-type: none"> • Useful in multi-activity, complex settings where contribution claims cannot be directly observed and measured • Can help evaluators test and build consensus for contribution from including stakeholders in assessing the strength of a causal contribution claim. 	<ul style="list-style-type: none"> • Needs experience and sensitive handling to be seen to work well • Reliability of assessment depends on selected probabilities • Not easy for some users to get to grips with (relies on formulas & probabilities). • Can be rigid and inflexible – testing parameters need to be anticipated early in methodology.
Contribution Analysis		
<p>CA has been around for some time but only recently gaining currency among evaluators. It focuses on assessing the likelihood of a contribution to an outcome or set of outcomes from the intervention. It provides for a seven-step, progressive process which maps a causal chain and adds additional evidence to that already available to look at how the contribution would have come about. It builds in allowance for other influencing factors and an embedded testing stage using ‘knowledge others’ (stakeholders) and can be used in diverse intervention settings.</p>	<ul style="list-style-type: none"> • Can be well fitted to evaluations where there is little scope for conventional counterfactual methods to assess contribution of an intervention. • Able to embed a theory of change, and use the progressive process to help critically review – or revise – it. 	<ul style="list-style-type: none"> • CA depends on the quality of the initial causal chain and the enhancements from added evidence • Key assessment stage is essentially subjective; lack of rigor may not give users used to quantification much confidence in the attribution • Not well suited to evaluation contexts where there is a lot of variation in implementation, or changes over time.
Contribution tracing		
<p>Not a variant of CA, but rooted in hybrid of Process Tracing (see below) and Bayesian updating using mixed (quali-quant) methods. Uses participatory methods to establish outcomes traces and using probability-based validity of contribution claims. Unlike CA, CT is a rigorous method guided by explicit criteria for data collection and measuring confidence and probability assessment to quantify the level of confidence in a particular contribution claim. Like CA, it builds in consultation with ‘critical friends’ and relevant stakeholders.</p>	<ul style="list-style-type: none"> • CT is a focussed methods using only evidence likely to increase or decrease confidence in specific contribution claim • Precise (guided) application contributes to the clarity and quality of the underpinning theory of change. 	<ul style="list-style-type: none"> • A rigorous method which needs systematic and careful handling especially of the undertaking schedule • Not suited to short duration intervention with insufficient time for ‘traces’ to be realistically observed

	<ul style="list-style-type: none"> Confidence in the analysis is enhanced by appropriate use of 'critical friends' during the testing phase 	<ul style="list-style-type: none"> Needs considerable time and care to explore alternative explanations.
Process tracing		
<p>A structured method centred on individual cases of change (which can be used in multiples) to test if a causal-effect expectation explains the outcome being assessed. Evaluators 'trace' outcome and implications which would be expected if the causal chain (theory of change) being tested were true. It can be combined with Bayesian updating (as above) to increase the rigour of causal claims. Uses various logical tests to help assess and demonstrate validity.</p>	<ul style="list-style-type: none"> Well suited where there is scope for case intensity and no counterfactual. Well suited to ex-post evaluation of a single case Can be used in multiple case situations (although risking complexity in explanation). 	<ul style="list-style-type: none"> Very high intensity method – not well suited to interventions with variation in application or non-stable contexts Needs high level of qualitative skills and used systematically and with rigour to prevent rater or inferential errors.
Qualitative Comparative Analysis		
<p>QCA is an established method which provides for systematic comparisons of outcome influences based on qualitative knowledge with some quantified testing to indicate reliability of assessments. It compares different aspects of an intervention effects set against contextual factors to identify various patterns and better understand the different characteristics (or combinations) linked to these.</p> <p>It is useful in complex settings where multiple influences need to be in place to achieve outcomes. Robustly applied it can identify success factors (and dis-enablers) and where these work in combination(s), and is well suited where there is expected to be considerable (eg geographic) variation in intervention effectiveness. It can work within a TBE by using the ToC to help anticipate factors of interest in transformation processes.</p>	<ul style="list-style-type: none"> Allows for both complex causation (combinations of factors) and multiple causes of an outcome to be accounted for usually in post hoc evaluation. QCA works best when data on all the cases of interest are available and the number of cases is neither too small nor too large, around ten to fifty cases. 	<ul style="list-style-type: none"> Not well suited where it needs larger numbers of cases for confidence in the analysis. Does not always provide for clear messages (eg I which cases represent more 'success' or 'failure' than others). Not a participatory methods so may present challenges in building confidence in findings, and allowing for unobserved (alternative) explanations.