Briefing note: QuIP in the context of existing evaluation approaches

A wide variety of other approaches to impact evaluation are in use including qualitative, quantitative, participatory and mixed methods and traditions (e.g. see http://betterevaluation.org/). The QuIP draws particularly on qualitative approaches, in the sense that it deals primarily with words rather than numbers, derived from open narrative text rather than responses to closed questions. Rather than drawing on its own distinctive body of theory it is also the product of a pragmatic, eclectic and iterative learning-by-doing approach to methodological development that borrows from several other approaches.

Evaluation methods:

Experiential Learning	QuIP
By experiential learning we mean what we learn as practitioners through first-hand experience of projects and direct communication with others directly involved with it. This is generally the first and most important source of evidence that informs organizational learning, and the counterpoint against which to assess evidence from other sources. It is also cheap. But is it enough? One problem is that personal familiarity with a project can restrict or bias the way you think about it.	The QuIP generates evidence with the added credibility of being collected in a more transparent way by researchers without prior understanding of the project, or an interest in its outcome. Their specialized expertise and professional reputation can also add to the quality and credibility of the evidence they provide.
Quantitative methods	**QuIP**
Quantitative methods, including randomized control trials, can generate precise estimates of the magnitude (and hence importance) of typical or average impacts, and these can be tested for statistical significance. They work best when project 'treatments', intended outcomes and the links between them that are relatively stable, clearly understood and easily quantified.	The QuIP aims to be useful in more complex and uncertain situations. The QuIP addresses both the exploratory and the confirmatory ends of the spectrum. It mostly only provides evidence of the nature of impact rather than its magnitude, but it offers a more detailed picture of how and why this varies within groups of respondents, as well as between them. It is also scalable: more interviews can be added as necessary to capture different experiences of intended beneficiaries. Sampling and questionnaire design can also be adjusted to focus a QuIP on more specific issues and/or sub-sets of intended beneficiaries. It also has the potential advantage of being both cheaper and more flexible than most forms of quantitative impact assessment.

Qualitative methods	QuIP
Many other forms of qualitative research can be incorporated into impact assessment, ranging from participant observation to process tracing. Confusion over the sheer variety of qualitative research methods and philosophies explains in part why they are not used more for impact assessment.	We make no claim that the QuIP methodology proposed here is better or worse than other qualitative approaches. However, we are strongly of the view that if it is to be used more frequently in impact assessment then the process of qualitative research, and the logic behind it, needs to be explained and presented more fully and openly. In the absence of a clear account of how qualitative research is conducted potential users are unable to distinguish between good and bad studies, and so end up using it less. Clearer and more thoroughly tested guidelines for using qualitative research in impact evaluation can also, we hope, help to reduce their cost and the time lag between commissioning them and obtaining results.

Contribution analysis

The QuIP has a strong affinity to Contribution Analysis (C) as described by Mayne (2012), as illustrated by the table below. Mayne (2012:273) also distinguishes between attribution ("… used to identify both with finding the cause of an effect and with estimating quantitatively how much of the effect is due to the intervention") and with contribution, that asks whether "… in light of the multiple factors influencing a result, has the intervention made a noticeable difference to an observed result and in what way?" Taking "observed results" to refer to changes measured through routine monitoring, the QuIP conforms to this definition of contribution. But as the basis for identification of causal chains it also conforms to the first part of Mayne's definition of attribution. Indeed, as an input into systems modelling and simulation it can also support some quantitative estimates of impact. By systematically reviewing evidence against project goals and theory the QuIP resonates with CA in aiming to serve a "confirmatory" purpose. But by asking blindfolded and relatively goal-free questions it also aims to serve as a more open-ended or "exploratory" reality check (Copestake, 2014).

Similarities:

- The importance of clear specification of the attribution claim, and the theory behind it.
- Gathering of appropriate evidence with which to confront the attribution/contribution story.
- Triangulation against other methods and through repeat rounds of data collection and analysis.

Contribution Analysis Steps	QuIP related activities
1. Set out the attribution problem to be addressed Staff of the implementing agency agree the cause-effect relationship to be assessed, including: • The nature and extent of the *contribution* it expects to make • Other potential key influencing factors	• Initial consultations to inform design of the QuIP study, ideally (but not necessarily) early in the life of intervention being evaluated.
Step 2: Develop a theory of change and risks to it The theory of change and results chain detail the assumptions and risks behind the expected causal chains, including external factors which may influence outcomes.	• A key input into design of a QuIP study, including linking it to change monitoring, sample selection and choice of output domains.
Step 3: Gather existing evidence on the theory of change • Evidence on results and activities (outputs and outcomes/impacts) • Evidence on validity of assumptions of theory of change • Evidence on other influencing factors	• Important to assessing the need, size and timing of a QuIP study. • Process data on how X and Y can also inform sample selection and disaggregated analysis of QuIP data. • QuIP data can also be triangulated against other evidence (feature 10).
Step 4: Assemble and assess the contribution story and challenges to it • Assess strength of causal links and patterns and credibility of theory of change overall • Identify any weaknesses in evidence	• QuIP reports set out details of multiple contribution stories and alternatives. • Transparency in coding and presentation of data facilitate identification of weaknesses.
Step 5: Seek out additional evidence • Review and update the theory of change, if needed, in the light of previous evidence • Gather additional evidence, for example from project staff, beneficiaries, synthesis reviews.	• Combine QuIP with other methods, including quantitative monitoring to inform micro-simulation. • Discuss findings and recommendations with staff and other stakeholders (un-blindfolding if appropriate). Revise accordingly.
Step 6: Revise and strengthen the contribution story	• Explore possible follow up data collection and analysis.

Realist evaluation

Similarities:

- Emphasis on "what works for whom in what circumstances"
- Complexity ontology: neither positivist not constructivist
- Relationships and choreography matter: organised distrust
- Emphasis on reviewing theories and contexts

Differences:

- QuIP as one tool for identifying multiple links between Contexts, Mechanisms and Outcomes ("CMO Configurations"). But interventions (X) are not the same as mechanisms.
- Choreography of who knows what when: e.g. double blind interviews versus open comparison of theories of change.

With its rallying cry of "what works for whom in what circumstances" (Pawson, 2013:15) there are many obvious points of affinity between the QuIP and Realist Evaluation (RE). At a philosophical level it also occupies an intermediate position between aspiring to contribute to the universal truths of positivist science and a constructivist denial of establishing any reality independently of the beholder. Truth is out there, but hidden behind perceptions. Our always imperfect attempt to groping towards it entails protracted confrontation of theory with multiple and often inconsistent sources of evidence, kept honest by openness and "organised distrust" (Pawson, 2013:18). This reflects the complexity of the world, which Pawson (2013: 33) depicts as encompassing variation in volitions, implementation, context, time, outcomes, rivalry and emergence ("VICTORE"). Managing this is only possible with the help of explanatory theory. This includes the theories of change that inform adaptation of QuIP field instruments and development of a sampling strategy at the design stage. It is also relevant to inductive data coding, analysis and interpretatiion. In contrast the emphasis with QuIP on blindfolding appears to depart from the more transparent process of reciprocal comparison of theories that inform at least some traditions of realist interviewing (Manzano, 2016).

The QuIP's openness to identifying multiple and distinct pathways linking X and Z to Y also fits well with RE's stress on distinguishing multiple and distinct CMO (context, mechanism, outcome) configurations, where X and Z can be equated with Contexts, Y can be linked to Outcomes, and the central evaluative task is to unmask the cognitive Mechanisms (in the heads of respondents) that link the two together. The potential for QuIP to be used as part of a mixed method approach also resonates with RE. Pawson (2008:19) suggests that "as a first approximation… mining mechanisms requires qualitative evidence, observing outcomes requires quantitative [data] and canvassing contexts requires comparative and sometimes historical data."

(p.19). Indeed one response to this is to classify QuIP as a "mechanism miner" that should always be part of a mixed evaluation strategy.

Feasibility and cost-effectiveness have also been important design criteria, as has been the ethical commitment to give effective voice to the concerns of the primary intended beneficiaries of development activities. However, it departs from many participatory approaches to evaluation in aiming primarily to generate evidence that is credible and useful to people not closely involved 'on the ground' in the activities being assessed. To date the QuIP has also not involved respondents directly in analysis and interpretation of the data as a mechanism for promoting empowerment (in contrast to other methods, including Sensemaking, Most Significant Change and PaDev, for example). This is something, however, a component that could be expanded in future (Copestake et al., 2016).

Process tracing

Similarities:

- Parity of effort in examining project theories of change and counter-explanations, including response bias.
- Focus on quality of evidence: smoking gun (=explicit) and hoop test (=implicit).
- How much evidence is enough? Potential to address this through Bayesian updating.

Differences:

- Process tracing seeks to confirm (or reject) predetermined theories and outcomes; QuIP can also be more exploratory.

As indicated, the QuIP can be viewed as one way of gathering additional evidence to test prior explanatory theory. Unprompted positive *explicit* evidence of attribution generated by the QuIP can be likened to "smoking gun" evidence of impact in a particular CMO configuration: significantly increasing confidence in the applicability of change theories behind the intervention. Positive *implicit* evidence is more akin to "hoop test" evidence, its presence is less conclusive, but its persistent absence would cast doubt on whether the intervention is working as expected (Punton and Welle, 2015). Viewed as a process of "Bayesian updating" (Befani and Stedman-Bryce, 2016) the accumulation of evidence can also potentially be used to judge whether the number of interviews and focus groups is sufficient. For example, if it is feared that rising profitability of cash crops might result in children being taken out of school to work on them, and if prior expectations of this are neutral, then a judgement can be made on how many negative results (i.e. that don't mention such an effect) would be sufficient to assuage the concern. In this and other instances, the role QuIP studies can play in process tracing is strongly enhanced by the strength of complementary evidence of change in key outcome variables, and this reinforces the argument for nesting use of the QuIP within a mixed method evaluation strategy.

The table below further compares QuIP with process tracing by relating it to ten "best practices" set out by Bennett and Checkel (2015:261). The QuIP also chimes with their argument for greater transparency with respect to the procedures used to collect and analyse evidence, and call for a "(partial) move away from internally generated practices to logically derived external standards" (p.266) without at the same time removing entirely a more exploratory "soaking and poking" of available evidence.

Process Tracing best practices	Relevance to the QUIP
1. Cast the net widely for alternative explanations.	The exploratory nature of the QUIP (use of open ended questioning and mitigation of potential pro-project bias) makes it open to a wide range of explanations, as does accommodation of multiple cases, and triangulation against evidence from focus groups, and other sources.
2. Be equally tough on the alternative explanations.	Evidence on project related and incidental drivers of change are collected and analysed in the same way.
3. Consider the potential bias of sources of evidence	Blinding aims reduce the dangers of intervention induced bias.
4. Take into account which explanations are most or least likely to explain a case.	Collection of data for multiple households (and through focus groups) helps to mitigate the risk of attaching too much weight to 'freak' instances.
5. Make a justifiable decision when to start.	Start linked to commencement of the intervention being evaluated and theories regarding its likely impact pathway.
6. Be relentless in gathering diverse and relevant evidence, but make a justifiable decision when to stop.	The number of cases assessed and process of selecting them can be adjusted to increase diversity of evidence, with the limit determined by accumulated experience of when diminishing marginal returns arise to increasing the number of interviews. Credibility is also enhanced through comparison with evidence of change in key variables obtained through quantitative monitoring
7. Combine process tracing with case comparisons when useful for the research goal and when feasible.	Comparison between households is integral to the approach. Standardization of the protocol also facilitates such comparisons. Sampling across complex contexts is a key issue in order to be able to address the counter-hypothesis that results are the product of selecting freak examples or outliers.
8. Be open to inductive insights.	The exploratory aspect of the QuIP (openness to respondents' own unprompted causal explanations) makes it open to these and to gaining insight into unforeseen consequences.
9. Use deduction to ask "if my explanation is true, what will be the specific process leading to the outcome?"	Interpretation of evidence is aided by triangulating it against steps in the prior theory of change for the project, and staged un-blinded triangulation whereby implementing staff can comment on findings.
10. Remember that conclusive process tracing is good, but not all process tracing is conclusive.	The methodology does not rule out being inconclusive about the relative contribution of different causal drivers identified.

While a strength of the QuIP is that it can be used as a stand-alone method without the need for a baseline, it can also utilised as part of larger and mixed method assessments. Five different ways of doing this are set out in the table below. These different models can also be combined.

Mixed method assessment incorporating the QuIP

Mix of methods	Rationale
1. Independent reality check or deep dive. Routine quantitative monitoring of key performance indicators (KPIs) to inform performance management. QuIP utilised selectively as an independent reality check.	Participation in operational activities provide staff with sufficient evidence to accurately interpret observed changes in KPIs most of the time. But QuIP can inform staff more remote from the field and also serve as a check against creeping biases and group think.
2. Combined process and impact evaluation. Use QuIP study before or alongside formal process evaluation. Doing it before enables evaluators to work back from impact to reviewing earlier steps in programme theory.	Process evaluation (by unblindfolded) researchers focuses on achievement of measurable outcomes. The QuIP focuses on shedding more accurate light on the difficult outcome to impact step in the theory of change
3. Parallel Q-squared impact assessment QuIP used in parallel with a quantitative impact assessment study (e.g. randomized control trial, or difference-in-difference studies) to facilitate interpretation of findings from the quantitative study.	Quantitative IA provides estimates of the magnitude of key inputs. QuIP facilitates interpretation of the causes behind observed changes and reasons for variation around average effects. Duplication also serves a quality assurance function – e.g. into whether observed changes are consistent.
4. Sequential Q-squared impact assessment QuIP used before or after a quantitative impact assessment study with the first being used to frame the focus and scope of the second.	Either QuIP is used to identify key causal for more precise estimation using quantitative methods, or as a follow-up to understand the causal processes behind observed correlations, ambiguous findings and unresolved issues.
5. System modelling and simulation QuIP used to identify key causal processes. This is then combined with quantitative data from other sources to inform modelling and simulation.	Simulation permits estimation of the magnitude of effects, and hence generate cost-benefit or cost-effectiveness calculations. Models may also be used to build more complex logic models and for sensitivity analysis (e.g. by distinguishing between impact for sub-categories of intended beneficiaries).

Tailoring methods to purpose:

Key evaluation questions	**RCT**	**Realist Evaluation**	**Process tracing**	**Contrib'n analysis**	***QuIP?***
How big was the overall change caused by the intervention?	5	1	1	1	1
What difference did it make to different sub-groups and why?	4	4	3	4	4
How and why did it make a difference?	2	5	5	5	5
What other factors affected these outcomes?	4	4	3	3	5
Which outcomes do different groups think most important?	1	2	2	1	3

References

Befani, B. and Stedman-Bryce, G. 2016. Process tracing and Bayesian updating for impact evaluation. *Evaluation*, 22(4), October.

Bennett, A. and Checkel, J. 2015. *Process tracing: from metaphor to analytic tool*. Cambridge: Cambridge University Press.

Copestake, J. 2014. Credible impact evaluation in complex contexts: Confirmatory and exploratory approaches. *Evaluation, 20*(4), 412-427. doi: 10.1177/1356389014550559

Copestake, J., Allan, C., van Bekkum, W., Belay, M., Goshu, T., Mvula, P., Remnant, F., Thomas, E., Zerahun, Z. 2016. *Managing relationships in qualitative impact evaluation to improve development outcomes: QuIP choreography as a case study*. Bath: Bath SDR Ltd, working paper. Accessed from qualitysocialimpact.org on 31 August. 2016.

Copestake, J. and Remnant, F. 2014. Assessing Rural Transformations: Piloting a Qualitative Impact Protocol in Malawi and Ethiopia. In: K. Roelen and L. Camfield, eds. *Mixed Methods Research in Poverty and Vulnerability: Sharing ideas and learning lessons.* Oxford: Routledge. Also available at: http://www.bath.ac.uk/cds/publications/bpd35.pdf

Gawande, A. 2008. Better: a surgeon's notes on performance. London: Profile books. See also The Positive Deviance initiative, basic field guide to the positive deviance approach. www.positivedeviance.org.

Manzano, A. 2016. The craft of interviewing in realist evaluation. *Evaluation*, 22(3):342-360, 2016.

Mayne J. 2012. Contribution analysis: coming of age? *Evaluation* 18(3):270-280.

Pawson, R. 2013. *The science of evaluation: a realist manifesto*. London: Sage.

Punton, M. and Welle, K (2015) Straws-in-the-wind, hoops and smoking guns: what can process tracing offer to impact evaluation? *Centre for Development Impact, Practice Paper No.10*. April.