QuIP

Qualitative Impact Assessment Protocol

This brief explains what the QuIP is for and outlines its approach to evidence collection, analysis and use. It is written primarily to inform potential commissioners of work using the QuIP. Two annexes elaborate on (a) sample selection and (b) how the QuIP compares with other approaches to qualitative impact evaluation.

1. Overview and background

Individuals and agencies who commit to actions with social and development goals need evidence about whether they are indeed achieving what they intended. Such actions may be referred to in many ways: as grants, investments, interventions, projects or programmes, for example (we use 'project' to refer to any of these). But in all cases the actors need evidence to help them decide whether to carry on, to expand or change what they are doing. They also need to inform those with whom they work, including people intended to benefit from the actions and those helping to finance it.

In diverse, complex and rapidly changing situations it is not obvious how best to obtain such evidence. This depends partly on why evidence is most needed. Is it primarily to demonstrate that past actions worked, or to find ways to improve on-going activities? Is it more important to quantify the magnitude of impact, or to explain why this varied from person-to-person or from place-to-place? How credible does the evidence have to be, and what level of expenditure on evidence generation can be justified? There are many different ways of answering these questions. Their strengths and weaknesses vary according to context, and no one method or approach outperforms all the others under all conditions.

This briefing introduces one method, referred to as the Qualitative Impact Protocol, or QuIP. Its main purpose is to serve as a reality check on whether the social effects of a planned activity or set of activities on intended beneficiaries is as expected, or whether it is having unintended consequences. It can reveal harmful effects, and throw light on other factors – some perhaps unexpected – that are affecting hoped-for changes. And it can also highlight variation in causes and effects across a group of intended beneficiaries. Generally, however, it is not so useful for revealing the *magnitude of changes*, and for this reason some people refer to it as a way of assessing impact contribution rather than attribution. However, it can usefully assist in estimating the magnitude of possible impacts when used in combination with other methods.

Put more formally, the QuIP aims to generate evidence on whether the casual links between 'project' activities (X) are contributing causally to a set of impact indicators (Y) under conditions of organised (i.e. not fully chaotic) complexity arising from the presence of interconnected, uncertain and hard-to-measure confounding factors (Z). In contrast to quantitative methods, the QuIP sets out to generate case-by-case evidence of impact based on narrative causal statements elicited directly from intended project beneficiaries without the need to interview a control group. Evidence of attribution is sought through respondents' own accounts of causal mechanisms linking X to Y alongside Z. This contrasts with methods that rely on statistical inference based on variable exposure to X. Such narrative data can usefully complement quantitative evidence of changes in X, Y and Z obtained through routine tracking or monitoring of key project indicators.

There are strong ethical grounds for asking people directly about the effect of actions intended to benefit them. Doing so can also contribute practically to learning, innovation and wider accountability. But doing so entails finding credible ways to address potential response biases. The QuIP does this by arranging for qualitative data collection to take place with as little reference as possible to the specific activity being evaluated, and by giving equal weight to all possible drivers of change in possible domains of impact (such as increased food security or household income) of interest. This is achieved by working, where possible, with field researchers who are completely independent of the organization responsible for the actions being evaluated. Indeed where possible field researchers are 'blindfolded' from knowing the identity of the organisation being evaluated, as well as details of what it is doing, and the theory of change behind its actions. Evidence collected from respondents in this way takes the form of narrative statements about causal drivers of change in selected areas of their life. Another researcher, the analyst (who is not blindfolded), then organises and analyses these using a standardised approach to coding which, working back from reported outcomes, highlights whether the reasons given for change explicitly or implicitly confirm or undermine the causal theory underpinning the intervention, or are completely incidental to it. Where available, this analysis can then be compared with data or project activities and observed changes, helping to build a more detailed picture of what has really changed and why.

Even if potential response bias is absent, another potential limitation of self-reported evidence of impact is that it is restricted to what respondents actually know and regard as most important. Use of the QuIP does not depend on believing that respondents are all knowing, but their experiences and opinions are valid and important. It can be used alongside other forms of evidence to identify important cognitive gaps between different actors. For example, if project staff and intended beneficiaries do have widely contrasting perceptions then it is likely to be useful to know this.

While the QuIP builds on many earlier ideas and experiments it was formally developed in its current form through a three year action research project between 2012 and 2015. This was referred to as the 'ART' (Assessing Rural Transformations) Project, led by researchers at the University of Bath, with research funding from the UK Government. The ART Project set out to design and pilot a credible way to assess the impact of development activities in the context of complex processes of rural transformation across Africa. It drew on various more established qualitative approaches, including contribution analysis, process tracing and realistic evaluation. Annex 2 elaborates on the relationship between the QuIP, these and other impact assessment methods. Pilot studies were carried out with researchers in Malawi and Ethiopia, and in collaboration with three non-government development organisations: Self Help Africa, Farm Africa, and Evidence for Development. The subject of these studies were projects that aimed to strengthen rural livelihoods and food security at the individual, household and community level in the context of both rapid commercialisation and climate change. Since then, the QuIP has been used in a range of other countries and contexts.

The ART project produced a comprehensive set of QuIP Guidelines which can be freely downloaded and are now available at www.bathsdr.org. This includes the original household level interview and neighbourhood level focus group questionnaires. The ART project trialled the QuIP on rural livelihoods projects, but the QuIP has since been applied to a range of other contexts, where household and local level impacts are important. Since the end of the ART Project, an independent non-profit research organisation has been established to continue developing and disseminating the QuIP. **Bath Social & Development Research** offers consultancy services to undertake QuIP studies as well as training in the approach. BSDR is also authorised by the University of Bath to support licensed use of the bespoke Excel spreadsheet for data analysis.

2. QuIP Methodology

Data Collection

The QuIP relies on a mixture of semi-structured interviews with individuals (at the household level) and focus group discussions (at the neighbourhood level) to assess impact based on self-reported attribution. The interviews are carried out by local independent field researchers who are informed as little as possible about the organisation and project being assessed. The purpose of this blindfolding is primarily to reduce potential for pro-project bias on the part of respondents, including their response to cues from the researchers. Individual and focus group respondents are asked a series of open-ended, non-project specific questions about any changes in their lives and livelihoods over a specified period of time, covering selected domains of well-being within which social effects are anticipated. The impact domains are decided upon depending on the type of project being implemented, for example the rural development projects assessed during the ART project included food production, cash income, cash spending, food consumption, intra-household relationships, household assets and external relationships. Most questions are open-ended, aiming to elicit respondents' own account of both what has changed in each domain and why. But discussion of drivers of change in each domain ends with closed questions to establish clearly the respondent's own views about how this domain of their life has changed overall during the specified time period. This helps to close each section of the interview. It also provides a useful snapshot of respondents' overall experience of change. The table below illustrates what this information looks like. The data can also be compared against both the narrative data and monitoring data from other sources.

Example of responses to closed questions (self-evaluation of change over a specified period in selected domains)

HH Code	Main respondent	Age of respondent	1. Food Production	2. Cash income	3. Cash Spending	4. Food consumption	5. Assets	6. Overall Wellbeing
TG1	Female	33	+	+	+	+	+	+
TG2	Male	38	-	-	-	+	+	+
TG3	Male	37	+	+	+	+	+	+
TG4	Female	52	+	-	-	=	-	+
TG5	Female	52	-	-	-	=	-	-
TG6	Female	40	-	=	+	+	+	+
TG7	Female	47	+	+	+	+	+	+

Field researchers conduct the interview in the local language in a conversational style, allowing plenty of time and encouraging a storyteller-listener rather than an interviewer-respondent relationship. To this end field researchers take notes on a paper pro-forma rather typing into a computer or tablet, which can be distracting. The notes are summarised and translated into English, with audio recordings used as a back-up wherever respondents consented to this. A comprehensive guide for field researchers (in English, French and Spanish) can be found in at www.bathsdr.org/resources.

Sample size and selection strategy is an important determinant of the scope for generalising from data collected using the QuIP. There are of course fixed costs associated with doing any study, and the guidelines are based on the judgement that anything less than employing two field researchers to spend five days in the field collecting data is likely to be hard to justify. On the other hand, there may well be a case for commissioning more than one study of this size – e.g. conducting two QuIP studies in contrasting geographical areas. Annex 1 elaborates on sampling and stratification options.

Data Analysis

A common issue with qualitative impact assessment is how to organise and make sense of large quantities of textual data, and to do so in a way that is transparent, so that generalisations drawn from it can be peer reviewed. The issue is most commonly dealt with by coding and summarising the data either 'deductively' using predetermined themes, or more 'inductively' by identifying repetitions and patterns through immersion in the data. Various software packages are available on the market to assist with the task, but the QuIP relies instead on use of preformatted Excel spreadsheets. This reflects a preference for flexibility, accessibility and keeping down costs. Where possible two or more levels of coding are used, employing standard routines to aid speed and transparency.

Unlike the field researchers, QuIP data analysts need to be fully briefed about details of the project. This is because their first coding task is to assess how the data relates to the project's theory of change by systematically identifying cause-and-effect statements embedded in it according to whether they (a) **explicitly** attribute impact to project activities, (b) make statements that are **implicitly** consistent with the project's theory of change, (c) refer to drivers of change that are **incidental** to project activities. These statements are also classified by outcome domain, and coded according to whether respondents described their effects as positive or negative. These fixed 'attribution codes' are detailed below.

Attribution codes for confirmatory analysis of project impact

Positive		**Negative**
1	Change explicitly attributed to project and project-linked activities	2
3	Stories confirming/questioning a mechanism by which the project aims to be achieving impact, but with no explicit reference to the project	4
5	Change attributed to any other forces that are not related to activities included in the commissioning agent's theory of change	6
7	Change not attributed to any specific cause	8

Summary tables indicating what text has been coded in this way provide a quick overview of the extent to which the data collectively validates or challenges the theory of change behind the project, illustrated by the tables below. The QuIP Excel spreadsheets produce these tables automatically, and frequency counts can also be exported as numerical data. These tables also make it easy for readers to refer back to the source text behind each coded response to find out more precisely to what each refers.

Illustrative summary table of positive drivers of change by attribution code and outcome domain

	1. Project explicit	3. Project implicit	5. Other	7. No cause
Food production	TG2 TG7 TG10 TG13 TG14 ***TGF1 TGF3 TGF4***	TG3 TG4 TG9 TG12 TG13 ***TGF2***	TG1 TG3 ***TGF3***	
Cash income	TG7 TG10 ***TGF1***	TG3 TG8 TG9 TG12 ***TGF2 TGF4***	TG1 TG3	
Cash spending	***TGF1***	TG3 TG7 TG10 TG12	TG1 TG3 TG6	***TGF4***
Food consumption	TG1 TG7 ***TGF1***	TG2 TG3 TG12 TG16 ***TGF2 TGF4***	TG9	TG6 TG10
Relationships		TG3, TG7 ***TGF1***	TG1 TG2 TG4 TG10 TG14 ***TGF3***	***TGF3***
Asset accumulation	***TGF4***	***TGF1 TGF2***		
External relationships	TG1 TG7 TG10 TG12 TG14 TG16 ***TGF1 TGF2 TGF4***	TG1 TG2 TG3 TG7 TG11 TG14 TG16 ***TGF1 TGF2 TGF4***	TG1 TG2 TG10 TG12 TG15 ***TGF4***	

TG refers to individual household codes. ***TGF*** refers to focus groups: TGF1 Younger women; TGF2 Older women; TGF3 Older men; TGF4 Younger men.

Illustrative summary table of negative drivers of change by attribution code and outcome domain

	2. Project explicit	4. Project implicit	6. Other	8. No cause
Food production		TG2 TG5 TG6 TG7 TG8 TG11 TG14 TG15 TG16	TG4 TG5 TG8 TG9 TG10 TG16	
Cash income		***TGF2 TGF4*** TG2 TG8 TG11 TG14 TG15 TG16 ***TGF1 TGF2 TGF4***	***TGF1*** TG4 TG13	
Cash spending		TG8 TG11 TG14 TG15 TG16 ***TGF2 TGF4***	TG2 TG3 TG4 TG5 TG9	
Food consumption		TG5 TG8 TG14 TG15 ***TGF2***		
Relationships			TG5 TG9 TG 10 TG11 ***TGF1 TGF4***	
Asset accumulation		***TGF2***		
External relationships	TG5 TG6	TG15	***TGF4***	

The second level of coding is more inductive, with analysts identifying and grouping together more specific causes or **drivers** of positive and negative change by outcome domain, whether attributable to the project or not. Once this task is done the QuIP Excel spreadsheet again produces summary tables, as illustrated below. These also make it easy to drill down into the source text behind each coded item. The data can also be exported in numerical form, and it can be further interrogated in other ways - to explore further the distribution of responses according to socio-economic characteristics of respondents, for example.

Illustrative summary table of drivers of positive change by cause and outcome domain

	Food Production	Cash Income	Cash expenditure	Food consumption	Relationships	Assets	Overall wellbeing
Supply of improved seeds, tools, irrigation and fertiliser by 'xyz' for production of fruit and vegetables	TG3, TG4, TG7, TG12 ***TGF1, TGF2, TGF3, TGF4***	TG3, TG7, TG10, TG12 ***TGF1, TGF3, TGF4***	TG7, TG10, TG12 ***TGF1***	TG1, TG2, TG3, TG7, TG12 ***TGF1, TGF2, TGF3, TGF4***		TG7 ***TGF4***	TG1, TG6, TG7 ***TGF1, TGF2, TG4***
Goat rearing; goats purchased using credit scheme provided by 'xyz' or given by 'xyz'	TG3 ***TGF1, TGF3, TGF4***	TG8, TG9, TG11 ***TGF1***				TG1, TG2, TG8, TG9, TG14, TG16 ***TGF1, TGF2, TGF3, TGF4***	TG1
Beekeeping	TG13 ***TGF4***					***TGF4***	TG1, TG7

Illustrative summary table of drivers of negative change by cause and outcome domain

	Food Production	Cash income	Cash expenditure	Food consumption	Relationships	Assets	Overall wellbeing
Increased prices			TG1, TG2, TG3, TG5, TG6	TG5			
Adverse weather - hail in August, then shortage of rain.	TG2, TG5, TG7, TG9, TG10, TG11 ***TGF1***	TG2, TG5, TG9, TG11	TG2, TG11	TG5			
Problems maintaining livestock (chickens & goats)	TG5, TG8	TG2, TG5, TG5, TG9, TG11, TG15				TG2, TG5, TG6, TG15	

More detailed analysis is possible if QuIP data can be supplemented with information from project staff about each respondent's precise involvement with the project (e.g. training, receipt of cash transfers or in-kind inputs). This then permits the coding to also reveal gaps in responses, or highlight areas where respondents have fared badly relative to that which might have been expected.

The third and fourth levels of coding identify selected **outcomes** and **impacts** of change in more detail than permitted by relying on the outcome domains used to structure interviewing. This in turn allows for more detailed analysis of different cause-outcome configurations. For example, drawing on the tables above it might be useful to explore in more detail precisely what additional assets were acquired by those who had benefitted from goat rearing, or which crops were affected by adverse weather. An example table (from a different project) showing the reported drivers and outcomes demonstrates how these causal links can be highlighted, and these can be clustered together and visualised in network diagrams where they have led to wider and longer term impacts.

DRIVERS	Increased hh income/ savings	Improved community social cohesion	Increased yield	Improved IYCF practices	Increased crop variety	Improved hygiene	Improved antenatal health	Improved food storage methods	Improved nutrition	Varied diet	Increased food security
Access to fertiliser/seeds/chemicals	5	-	18	2	27	-	1	-	1	9	6
Provision of information on nutrition	-	-	-	26	-	-	17	-	17	2	-
Government (social cash transfer)	17	1	2	-	2	-	-	-	-	-	-
WASH information	1	-	-	-	-	29	-	-	-	-	-
Provision of livestock/poultry	6	-	1	-	-	-	-	-	-	-	-
Joined savings group	10	3	-	-	-	-	-	-	-	-	-
Training on post-harvest processing and handling	-	-	-	-	-	-	-	20	-	-	-
Joined cooperative/s	2	8	2	-	-	-	-	-	-	-	-
Providing information on better farming methods	1	-	6	1	1	-	-	-	-	1	1
More land cultivated	3	-	8	1	1	-	1	-	-	-	2
Increase in attendance at community meetings/training	-	11	-	-	-	-	-	-	1	-	-

Numbers refer to instances where narrative evidence has been coded with both the specified driver and outcome, indicating a causal relationship

For more information see http://bathsdr.org/about-the-quip/coding-and-visualisation/

Use of QuIP data and analysis

The data analysis described above can be adapted and taken further in numerous ways. The summary tables are typically incorporated into a written report that also pulls out quotes from the source narrative data to illustrate and elaborate on key findings. But use of findings does not have to rely on written outputs. For example, if trained staff from within the commissioning organisation do the coding themselves then internal learning starts even before the analysis is complete. An interim report can also be used to structure feedback meetings with project staff, individual respondents and other stakeholders. These are a form of triangulation, enabling those attending to challenge, corroborate and complement findings, thereby both serving a quality assurance function and deepening understanding of what changes took place for whom, how and why. Such debriefing meetings can be further enriched by 'un-blindfolding' field researchers thereby enabling them to offer their own interpretation of the findings, drawing on what they wrote down, their direct field observations and wider experience.

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.

Table 1: 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 un-blindfolded) 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. randomised control trial, or difference-in-difference study) to facilitate interpretation of findings.	Quantitative IA provides estimates of the magnitude of key impacts. 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 drivers 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).

Annex 1. Sample selection

There is no universal best practice method for sample selection for a QuIP study, as it depends upon many contextual factors. The most important of these are (a) the main purpose of the study, (b) availability of relevant data about variation in the characteristics of expected gainers and losers from the 'project' or intervention being assessed, (c) availability of relevant data about variation in their exposure to project activities, (d) time and resource constraints. This section briefly explores these factors, and then outlines the sequence of sampling decisions and actions needed prior to starting data collection.

(a) Main purpose of the study

Deciding who to interview, how many people to interview, and how best to select them requires clarity about what information is being sought, by whom and why. Neglecting this not only leads to poor practice but also misunderstanding about the quality of a study. For example, sample bias is not a problem for a QuIP study that deliberately set out to identify drivers of successful outcomes by interviewing people thought to have been particularly successful, or what Gawande (2008) and others refer to as "positive deviants." Deliberately selective (hence biased) sampling is in this instance fit for purpose.

More generally, differences in sampling strategy arise from whether the priority is to confirm and quantify the overall impact of a completed project on a defined population in relation to a predetermined set of measurable indicators, or to identify and explore what is happening in a more open-ended way – to improve implementation of an on-going project, for example. The QuIP is a relatively open-ended approach. Its primary purpose is to gather evidence of causal processes at play, not to quantify them. Deciding on the number of interviews and focus groups to conduct depends less on reducing sample bias than on assessing at what point the extra insight into the range of possible causal processes influencing outcomes gained from collecting more data no longer justifies the extra cost. To do this formally would not entail estimating statistical sampling errors but a Bayesian process of assigning confidence parameters to prior expectations and assessing how these change with each extra case or observation (see Befani and Stedman-Pryce, 2016). In contrast, if the primary purpose is to quantify specific causal effects then there are two options. The first is to use an appropriate experimental or quasi-experimental approach instead. The second is use simulation models, drawing on QuIP data to identify the main causal factors, and quantitative data from other sources to calibrate their magnitude.

(b) Contextual variation

Random selection of respondents across the entire population affected by the project is a good starting point for thinking about sampling for a QuIP study, but there are also good reasons for making adjustments for it. If we expect causal processes to be different for different sub-groups, and we have data that enables us to identify those sub-groups prior to sample selection then there is a case for stratified random sampling. For example, we might choose to ensure the QuIP study includes a minimum quota of people living in urban and rural areas. Stratification of the sample on these grounds is an art not a science that depends on prior thinking about what observable contextual factors are most likely to be a source of variation in project outcomes. Stratification choices also depend on what contextual data available. For example, if respondents are being selected from a baseline survey then it might be useful to stratify on the basis of baseline income or wealth indicators. And if respondents can be selected from longitudinal survey data then it might be interesting to stratify according to variation in respondents' experience of change in key indicators such as income or wealth. Combining these two examples a powerful design is a quota sample of four groups: richer and improving; richer but declining; poorer but improving; poorer and getting worse.

(c) Exposure or 'treatment' variation

This refers to variation in how project activities touch different people, including those who are direct beneficiaries of different packages of goods and services and those who may only be affected indirectly. If data is available on variation in who directly received what and when, and it is expected that these differences will have different causal effects, then there is a case for stratifying the sample to ensure it reflects the full range of such exposure. This is particularly the case if one purpose of the study is to aid decisions about which of a range of project activities or packages to expand or to stop. Impact assessment using the QuIP does not require a control group of people completely unaffected by the project. There may nevertheless be an argument for interviewing some people unaffected by the project, but similar to those affected by it in order to explore whether they come up with a different set of drivers of change.

(d) Time and resource constraints

In addition to stratifying according to contextual and exposure variation a third reason for departing from pure randomisation in sample selection is to cluster respondents geographically in order to reduce the time and cost of data collection. One way to do this is to adopt two stage random sampling, with the first stage based on geographical units (e.g. villages, neighbourhoods, districts or census areas). There is often a strong case for using contextual information (e.g. about agro-ecological zones) to purposefully select or at least stratify area selection. Ultimately, budget constraints may also limit the total number of interviews and focus groups that the QUIP study can cover. There may also be a case for staggering studies – i.e. conducting two smaller studies a few months apart rather than doing a single larger study. This can help to build understanding of project impact lags, pathways and cumulative processes, as well as those of other drivers of change. Sampling strategy for repeat studies can also be informed by lessons from earlier studies. Again the principle here is that credibility of findings builds incrementally with the addition of each extra piece of evidence.

Annex 2. What the QuIP adds to 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.

Realist Enquiry

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 interpretation. 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, however, a component that could be expanded in future (Copestake et al., 2016).

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).

Table 2: QuIP and Contribution Analysis compared

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.

Process tracing

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 overleaf 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.

Table 3. QuIP and Process Tracing compared

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	Blindfolding aims to 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-blindfolded 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.

Outcome Harvesting

Outcome Harvesting (OH) provides another interesting point of comparison: more so indeed than its name (cleverly dodging the words impact and evaluation) implies. It can be defined as "*an evaluation approach that does not measure progress towards predetermined outcomes, but rather collects evidence of what has been achieved, and works backward to determine whether and how the project or intervention contributed to the change.*" (UNDP, 2013, p.5)

This brief comparison draws primarily on a summary of the approach produced for the Ford Foundation in 2012 by Ricardo Wilson-Grau (the main originator of OH) and Heather Britt. A striking similarity with QuIP is the emphasis on garnering useful evidence of change and its causal drivers by **working back from outcomes** to activities of the commissioning organisation (referred to as the "change agent") rather than forward from the activities that it wishes to assess. Second, they also emphasise the usefulness of this approach to assessing outcomes in complex contexts where many factors and combination of factors may lead to many outcomes (positive and negative, anticipated and unanticipated), and where relations of cause and effect are not fully understood. Third, and linked to this, OH shares with QuIP an emphasis on the usefulness of gathering credible evidence of contribution, without necessarily being able to estimate precisely how much of a given outcome can be attributed to a specified activity. Indeed implicit in both approaches is recognition that aspirations to measure change and attribute outcomes too precisely may even be an obstacle to a broader and more reliable assessment of causal processes associated with the activities being assessed.

These commonalities with QuIP, allowing with more detailed differences, can be elaborated by looking in turn at the six iterative steps of Outcome Harvesting: (1) Design, (2) Identification and drafting of outcome descriptions, (3) Engagement with change agents in finalising these, (4) Substantiating outcome descriptions through consultation with independent agents, (5) Analysing, interpreting and making sense of the evidence, (6) Engagement with potential users of the findings.

Table 4. QuIP and Outcome Harvesting compared

Outcome harvesting steps	Comparison with the QUIP
1. Design. Identify useful questions and information to be collected through discussion with the change agent. This includes identifying key 'social actors' affected by the actions of the change agent. Key questions include the following. What happened? Who did it? How do we know? Why is it important?	QuIP starts with dialogue between the commissioner and lead researcher, including identification of the activities to be assessed, intended beneficiaries, which of them to interview and what potential outcomes ('domains') there should cover.
2. Gather data and draft outcome description. **3. Engage change agents in formulating the outcome description.** This entails "gleaning" data from readily available sources and organising it into a coherent set of outcomes and factors contributing to them. Chosen outcomes for description should be specific and realistic (e.g. about time lags, possible causal links), verifiable and relevant. The level of confidentiality should also be discussed.	QuIP requires that the lead researcher elicits from the implementing agency a clear 'theory of change', including as much detail as possible about what activities selected interviewees participated in and when. Discussion also covers how to approach interviewees, and how to frame discussions with them, including the choreography of blindfolding and un-blindfolding activities once data collection and analysis is complete (see below).
4. Substantiate. This entails obtaining the view of independent individuals ('substantiators') about the selected outcomes and how they were achieved. Their feedback affirms or challenges the credibility of the initial outcome descriptions. Substantiators may include key informants and/or panels of experts.	QuIP does the same but in a more prescriptive way through purposive sampling and interviewing of intended beneficiaries of the activities being assessed. QuIP seeks to enhance the credibility of this evidence through blindfolding. The more fluid and open approach adopted by OH appears closer both to process tracing and realist evaluation.
5. Analyse, interpret and make sense of the evidence. This is more straightforward for assessment of the contribution of one project by one change agent within a single period. But generalisations may also be sought for multiple activities and agents over multiple time periods.	QuIP focuses on the simple case, but offers a more systematic approach to coding and analysis of multiple sources of evidence. Being more prescriptive it can interpret findings more rapidly and transparently. Clear and succinct visualisation of findings is also critical.
6. Engagement with potential users of the findings. While there is a strong emphasis on generating useful evidence it is also recognised that it is rarely possible to make specific recommendations for action, as these are likely to be informed by other sources of information and operational factors not addressed by OH, particularly in complex and rapidly changing contexts.	QuIP also emphasises the importance of active engagement, beyond presentation of a final report. Opportunities arise to stimulate constructive encounters between change agents and other social actors through 'un-blindfolding' meetings (between field researchers, commissioners, operational staff and interview respondents) to discuss findings and their implications for action.

Overall, this brief comparison suggests that the values and philosophy underpinning Outcome Harvesting and QuIP are very similar. In aspiring to produce evidence that is credible and useful to actors in complex contexts both implicitly counsel against pursuit of universal truths and perfectionism (including spurious precision, or what Mansky calls "incredulous certitude"). Both also recognise the limitations of having to rely on the cooperation and perception of stakeholders in any change process, but also appreciate the ethical as well as practical benefits of eliciting and comparing multiple perspectives. Both distinguish between evidence of change ('outcomes') and evidence of drivers of change, and favour starting with the first and working back to the second.

There are also significant differences. While OH is more detailed and prescriptive than Outcome Mapping (see footnote 13 of UNDP, 2013) it is significantly broader in scope than QuIP – e.g. in addressing recurring monitoring needs alongside the need to assess impact of specific interventions. QuIP is also more narrowly focused on securing the feedback of intended beneficiaries, in a way that it more transparent and open to auditing by third parties. OH, in contrast (and like process tracing) appears more tailored to assessing individual efforts: e.g. in advocacy, campaigning and policy engagement.

Overall, the key point is perhaps that they are mutually affirming approaches that belong to a broad family of more qualitative and interpretive approaches to assessing change. For all the confusion of terminology and acronyms there is much to be gained from the existence of a plurality of approaches. Attempts to list, review and classify different approaches more systematically for different fields can be useful, but if we accept the benefits of practice that is attuned to diverse, complex and evolving needs then we should neither expect nor hope that any overarching review will ever be definitive.

References

Befani, B with M O'Donnell (2017) Choosing appropriate evaluation methods: a tool for assessment and selection. London, Bond. https://www.bond.org.uk/resources/evaluation-methods-tool

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.

Spaggiari, L (2016) Guidelines on outcomes management for investors. Guidelines for integrating outcomes measurement, analysis and reporting into operations of inclusive finance investors, asset managers and other funders. European Microfinance Platform and Social Performance Task Force. https://sptf.info/images/Guidelines-on-Outcomes-Management-for-Investors.pdf

UNDP (2013). Discussion paper. Innovations in monitoring and evaluating results. November. Edited by Jennifer Colville. [Jennifer.colville@undp.org]

Wilson-Grau, R. & H. Britt (2013) Outcome harvesting. Produced for the MENA Office of the Ford Foundation. May 2012 (revised November 2013).