This paper was presented at "Towards Carnegie III", a conference held at the University of Cape Town from 3 to 7 September 2012.

Ask them!
Better measures to manage performance in poverty programs

**Andre Proctor and David Bonbright of
Keystone Accountability (www.keystoneaccountability.org)[1]**

'…being attentive along the journey is as important as, and critical to, arriving at a destination.'
Michael Quinn Patton

'At its heart, Keystone's constituent voice methodology promises to shift the very power to define 'success', and to declare when it has been achieved, into the hands of those that development organizations claim to serve. This shift toward greater agency is essential if we are to make development more effective.' ***Raymond C. Offenheiser, President, Oxfam America***

Introduction

It is surprising that in a major conference on poverty, such as this one, there is so little discussion on practical approaches and methods for learning about **how well** we do what we do and particularly how well we listen and respond to those in whose name we work. Keystone believes that this is a seriously neglected aspect of most poverty reduction programs.

But this is not a call for the kind of quasi-experimental quantitative impact evaluation (IE) that so many see as the 'gold standard' for proving the impact of poverty programs. Nor is it a call for Results Based Management (RBM) or the rigid linear Logical Framework Analysis (LFA) that many funders require that their grantees use to design, plan and report on their work. Keystone locates itself firmly among those evaluation and learning practitioners who recognise that developmental interventions usually involve complex change processes, over which organizations have limited influence or control[2]. We recognise that development does not happen in neat one, two or even five year periods with a clear beginning, middle and end. We believe that performance measurement needs to concentrate on improving rather than on proving, on dialogic learning rather than accounting, and on creating knowledge rather than claiming credit.

---

[1] Contact Andre Proctor in Cape Town at andre@keystoneaccountability.org .
Keystone Accountability is made up of three independent, self-governing non-profit organizations registered in England the USA and South Africa. It was formed in 2004 with the aim of exploring practical ways of bringing the voices of those intended to benefit from developmental programs into the design, implementation, performance measurement and improvement of those programs. It has partnered with leading international and local development agencies (including USAID, IFC and WORLD BANK), philanthropic foundations (including Ford, Gates and Rockefeller Foundations), international NGOs (including Oxfam, CARE and HIVOS) to pioneer new approaches to measurement, accountability and learning for developmental outcomes.

[2] For a short summary that sets out where Keystone stands in the field of evaluation see **Improving Impact Evaluation for Development**, by David Bonbright, Fred Carden, Sarah Earl, Sanjeev Khagram, Nancy MacPherson, Zenda Ofir, and Patricia Rogers, available at http://idl-bnc.idrc.ca/dspace/handle/10625/40519 .

Most importantly, we recognise that in most effective development interventions there is, as Terry Smutylo and Sarah Earl of the IDRC have put it, 'the need for endogenous organizations and communities to take ownership of program components. Local ownership requires the devolution of planning, decision making, and other elements from external actors to internal actors'[3].

This paper does not attempt a comprehensive analysis of measurement and performance management in poverty reduction programs and interventions. Rather, it offers a *practitioner's perspective* on a particular approach and method for generating, making sense of, and learning from performance and impact *data* that is grounded in empirically valid feedback from the intervention's primary constituents[4] – those intended to directly benefit.

We call the approach **Constituent Voice** (CV). CV is an empirically rigorous and systematic way of listening to and learning from our most important constituents – those in whose name we do our work – and then using this knowledge to improve the relationships and the effectiveness of programs.

CV does not replace traditional monitoring of a program's operations nor does it do away with the need for formal evaluation. But it does assert that systematically listening to our constituents provides an important – and too often neglected –source of real-time knowledge and insight that we can use to manage performance, strengthen relationships and improve outcomes.

Feedback data is an early indicator of the health of an intervention and of the changes taking place – planned or unplanned. It can be triangulated with other evidence of results (including outcome measures and operational monitoring information) to enrich understanding of what is happening now. Dialogue about its findings generates learning and enables course correction and innovation. It is a reliable predictor of future outcomes. And, over time, the accumulated evidence can reveal patterns – the ebb and flow of an intervention's impacts – that help understand the nature and process of change.
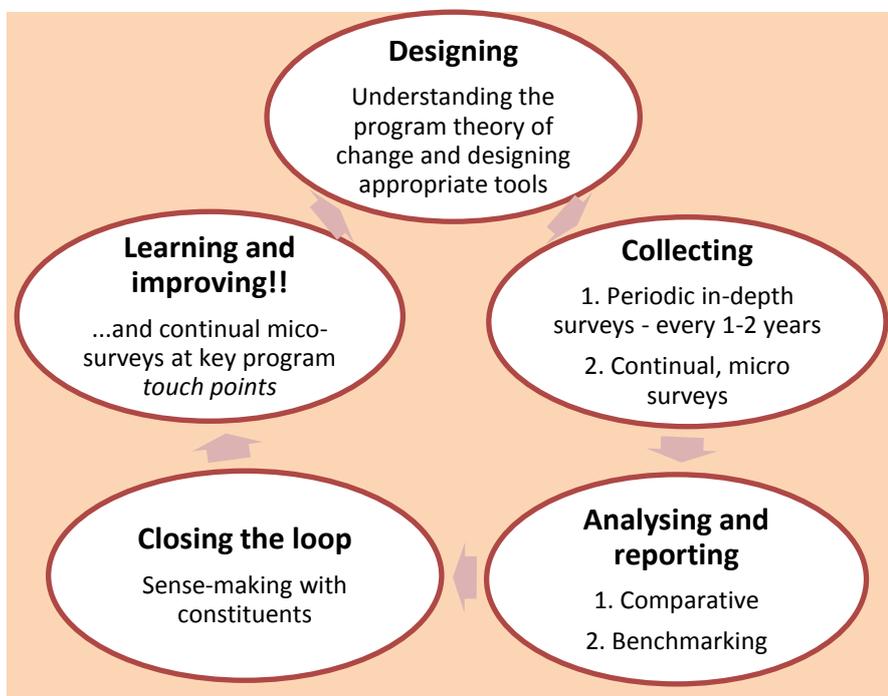
There are many different ways of generating CV, but over the last two years Keystone has worked with many different partners and arrived at a core methodology that consists of five linked elements or stages:

- Designing

- Collecting information

- Analysing and reporting

- Closing the loop

- Learning and improving

The paper will discuss each of these in turn.

---

[3] Terry Smutylo and Sarah Earle, **Outcome Mapping**, IDRC, Toronto 2002.
[4] Broadly there are three main constituents of most development interventions: those who fund it (funders), those who implement it (implementers – working alone or in partnership with others), and those intended to benefit (who we refer to as the intervention's 'primary constituents' in preference to the term 'beneficiary' emphasising agency rather than passive receiving of dispensed benefits).

Our early experience with CV convinces us that CV measures and accompanying practice of public reporting and dialogue are a powerful way of enabling real citizen engagement in the programs and interventions that claim to serve them at a local level. It is light-touch and cost-effective, and this makes it practical, replicable and scalable. By quantifying and aggregating qualitative feedback data we believe that it is possible to establish credible performance benchmarks for similar programs. CV enables real time local, regional, provincial and national comparison and learning about what interventions are working well, what lessons can be learned and where management interventions are required before problems become too serious. Most importantly, it turns subjects into actors.

We keep the paper deliberately short and focused on the principles behind CV and our experiences in the field designing and implementing CV measures with a range of development partners. The purpose is to stimulate new and creative thinking about measurement and performance management in ways that actually enhance intended outcomes: that foster transparency and accountability to citizens, that balance the power relations in management decision-making and that encourage real-time learning and improving through dialogue.

Constituent Voice is a work in progress. This paper is also an invitation to collaboration. Most of our experience to date has been in countries outside South Africa, but we have a number of local learning partnerships in the making. We would welcome more.

> **Constituency Voice and Ugandan Public Health Services**
>
> An important recent study has illustrated how constituency voice can improve community health. In this case, a government-supported social accountability project in Uganda involved citizens in a rural district organizing to monitor, assess and report on the quality of the local clinics' care.
>
> The study established that the citizen report cards on the clinics caused several significant benefits. Waiting times decreased. Doctor and nurse absenteeism plummeted. Clinics got cleaner and fewer drugs were stolen. Clinics were used more; forty to fifty percent more children were vaccinated.
>
> Most importantly, one-third fewer children died under the age of five. In the region served by the clinics, an area of 55,000 households, an estimated 550 lives were saved.
>
> None of these improvements were realized at a number of control clinics where citizen report cards were not implemented.

Step 1 - Designing

Each application of Constituent Voice is based on a thorough understanding of context, constituents and the program's theory of change.

Survey instruments must be based on well-validated survey questions and should be informed by input from many stakeholders. As more surveys are conducted and lessons learned, so the bank of well-constructed and meaningful questions grows. This makes the design process considerably easier as there is a wider range of well-tested questions to select from. Our experience to date suggests that there are five broad areas to cover in most development contexts:

1. Demographic questions for cross-tabulation and comparative analysis.
2. Constituent experience of specific operations and activities – the relevance and quality of services.
3. Perceptions of performance and effectiveness.
4. Perceptions of outcomes and impact – actual changes in constituents' lives, whether intended or unintended.
5. A relationship scorecard – a set of 7 questions on key relationship dimensions (see p. 4 below).

Step 2 - Collecting data

CV metrics make use of innovative survey techniques and information technologies to generate accurate, timely and actionable performance data at all levels of management.

Keystone has made a careful study of customer staisfaction measurement in the business world and combined its insights with our own deep experience of participatory learning to develop a new and effective mix of data gathering methods. Keystone's core hypothesis is that just as customer satisfaction has proven to be a reliable predictor of business success, so even more so can CV, when properly measured, be the best possible predictor of development outcomes.

We have identified two main methods of data collection.

1. Periodic comprehensive and anonymous feedback surveys conducted by an independent agency.

At the level of specific interventions (where most of the data is generated), carefully planned independent and anonymous surveys of a representative sample of beneficiaries provide robust perceptual data on the relevance and quality of services, the self-confidence of constituents, the quality of relationships as well as perceived progress towards outcomes.

The importance of independently collected anonymous feedback was demonstrated in a recent experiment in Tanzania where Keystone was surveying households in the villages affected by a sustainable forestry project. We created two research teams. One clearly identified itself as representing the forestry company. The other presented itself as an independent research firm contracted by the company. The feedback collected by the company researchers was consistently around 20% more positive than the feedback collected by the independent team.

These surveys should ideally take place every 2-3 years. They ask a number of questions derived from the intervention's theory of change, and can provide empirically valid data on three main aspects of an intervention's performance:
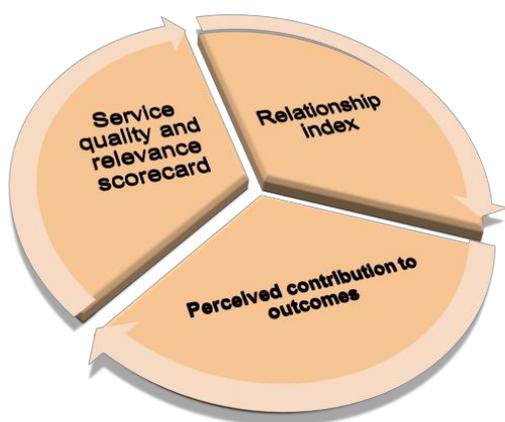
- Service quality and the relevance of activities to constituents' needs.
- The quality of relationships. Separate questions cover different relationship dimensions such as Confidence (in the knowledge and capabilities of program actors); Integrity (perceptions of honesty, commitment and fairness); Voice (do they feel respected, that their voice is encouraged and is heard, and that they belong and play a meaningful role in the intervention); Empowerment (do they feel the intervention is helping them stand on their own feet); Readiness (their level of confidence and commitment to invest their own time and energy); and overall satisfaction (how strongly would they recommend the program and/or its activities to someone else like themselves).

- Perceived contribution to outcomes. What are the changes that constituents themselves see in their lives and to what extent do they attribute these to the intervention.

The comprehensive surveys include both quantifiable rating questions (using a 0-10 Likert scale) and open-ended questions. Selected demographic questions enable disaggregation and comparison of responses by age, sex, location, level of education, employment and/or any other category relevant to the intervention.

Quantification enables comparison of program performance over time and across interventions or across regions to help identify which are working better from the beneficiary perspective and why. The open questions enable unstructured insights to emerge, and can also be analysed and quantified using standard statistical techniques.

Keystone has recently conducted comparative surveys of over 400 coffee farmers in Nicaragua on their perceptions of an IFC supported program to improve coffee production and marketing systems. We have surveyed the grantees of a number of East and Southern African development grant makers as well as the southern partners of approximately 30 leading international development NGOs. In most cases, clients have reported that the surveys generated valuable insights and that they have made changes as a result.

2. On-going micro-surveys at key program *touch points*.

The comprehensive survey data is complemented by data from on-going micro surveys of constituents. Micro surveys tell you what is happening as you implement a program or make changes as a result of the feedback you have received. They give a picture of the ebb and flow of the intervention. They can act as an early warning system and they allow you to respond to challenges before they become threats.

Micro surveys are very short surveys ranging from a single question to five or six questions. They are conducted at significant moments in the course of a project, or at key *touch points* e.g. after a training session, at a meeting or an event. They provide a light-touch, continual 'finger on the pulse' of constituents' experience and perceptions of the project or program. The data can be fed into, and compared with data from the comprehensive surveys.

The principle of independent feedback does not have to be expensive. In a few of our pilots, Keystone is training and using community volunteers, students, and other local data collectors to impart rigour to data collection by guaranteeing anonymity and confidentiality of respondents.

Micro survey questions can ask for a 'single button' response or contain one or two demographic questions and closed and open questions. For example, a question such as: 'On a scale of 0-10, how strongly would you recommend this training workshop to a friend or neighbour?' could be followed by a request to please give reasons for their rating or mention any things they liked or disliked about the training.

Micro surveys are ideally suited to automated data gathering technologies such as SMS and handheld devices which have the potential to significantly reduce costs as well as improve timeliness. They create a continuous stream of real-time performance data on the constituent experience of activities and outputs, relationship quality and impacts of the intervention on livelihoods that allows organizations to track progress against outcomes, monitor relationships and triangulate with self-reported data on activities and outputs.

In other words, while the comprehensive surveys provide detailed periodic **snapshots**, the continuous feedback allows you to play the **movie** from the perspective of primary constituents and other stakeholders against key indicators.

Step 3 - Analysing and communicating the data

Quantified perceptual data can be analysed using standard statistical methods to give reliable insights in the perceptions of different groups of constituents.

Net Promoter Analysis

To analyse constituent satisfaction with different elements of a development intervention, Keystone has adapted a powerful and proven methodology from the customer satisfaction industry called Net Promoter Analysis (NPA). NPA classifies respondents into promoters, passives and detractors and calculates a single Net Promoter Score (NP Score).

NPA is disarmingly simple, but hundreds of the world's leading corporations use it. It has proven to be a reliable measure of customer loyalty, and a powerful lever for positive organizational motivation and change. A score of 9 or 10 is classified as a 'promoter', 7 or 8 as 'passive' and 6 or below as a 'detractor'. Our experience to date suggests that it is a much more useful way of presenting data than conventional statistical tools of means, medians and quartiles.

Figure 11   Provision of benefits and desire to work for [...]
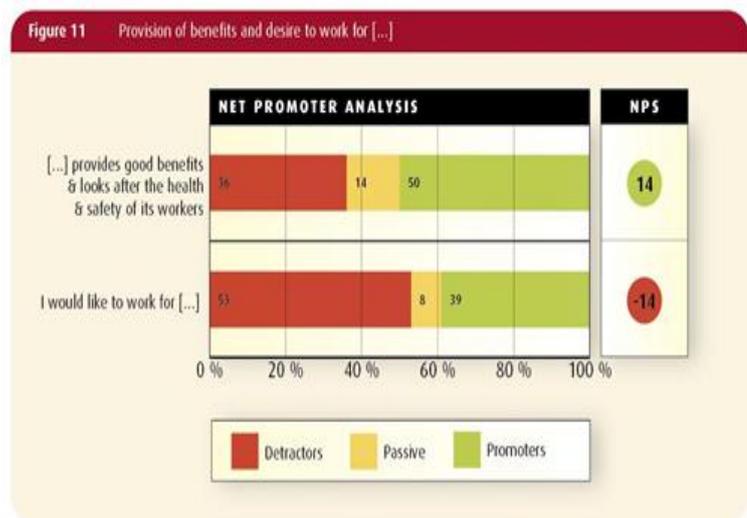
Scores for a number of related questions on a particular area of performance (such as the relevance and utility of activities), or from a number of short touch point surveys over a year can be aggregated into performance indexes for comparative analysis.

Data presented in this simple but highly effective way enables managers at all levels to utilize the data more effectively. It can be discussed and analysed within the program team or organization and used to identify areas for more in-depth investigation.
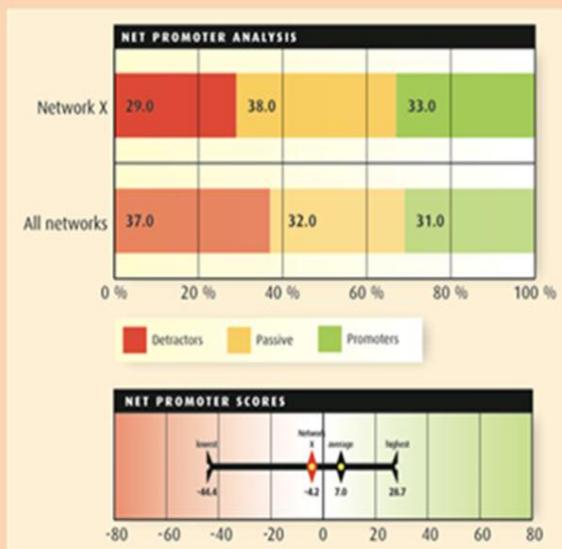
Benchmarking
Quantification of qualitative feedback data enables us to create performance benchmarks – not on the basis of technical 'ratings' by external inspectors, but on genuine internal feedback using common questions.

This approach to presenting bills to customers of a public utility in California produced a first ever dramatic reduction in energy use. Comparison is the key to getting people and organizations to act on metrics.

Tool 3: Service Quality

Published benchmarks and comparison with other similar interventions enables organizations to understand better what 'good' is for that kind of work. Organizations can compare their ratings with how other organizations are rated by their constituents – and gain a better sense of what is achievable and how they are performing relative to others like themselves. Benchmarks drive utilization of evaluative data – a strategy that the business world has perfected to a fine art. It is long past time we put this proven technique to work to solve important societal problems.

In business, customer ratings of goods and services are now very public, and are used by consumers in making their product choices. Companies cannot ignore what their customers are saying and put in a great deal of effort to gain positive reviews. Imagine what would happen if social purpose organizations knew that their potential funders were able to review independently collected constituent feedback on their work as well as on the work of other organizations like them. Public benchmarking could be a powerful driver of performance and greater accountability to constituents.

One example from Keystone's experience is found in our **Comparative Partner Feedback Surveys** for international development NGOs. Participants include CARE International, Oxfam, Save the Children, Concern and about twenty seven others. Keystone has asked thousands of developing country partners of these northern NGOs the same set of questions, and now has a central database with all the feedback data collected over two years.

This allows any northern NGOs to benchmark their feedback against all, or a selection of other NGOs like themselves. This not only gives us an idea of how they individually perform from the perspective of their southern partners, but also to draw general conclusions about the performance of northern NGOs as a sector, and explore ways of improving practice across the sector. It gives each an indication of how well they are performing, and the motivation to do better.

Step 4 - Closing the loop: sense-making with constituents

The real power of NPA and benchmarking lies in the possibilities it offers for reporting back to constituents and facilitating collective sense-making and dialogue for learning and improving. This fosters the kind of transparency and accountability to constituents that builds confidence and trust and enhances the credibility of the program among constituents.

When organizations report publicly in this way a new kind of learning conversation becomes possible between constituents and stakeholders. Even poor and illiterate communities are able to understand and make sense of feedback data. Our early experience of this kind of data-based dialogue is that it has led to deepened insights, new mutual commitments, improved program activities and strengthened relationships.

There are many different ways of reporting back and discussing feedback. Some clients have preferred independently facilitated meetings, others have organised a series of internal dialogues themselves. Where constituents are local, face to face meetings are possible. Where they are widely dispersed, a mix of print and other interactive media can be used. In two recent examples, Keystone facilitated report back dialogues with grantees of a major international funder that we had surveyed on their experience and perceptions of that funder. In one of these, after going through the reports privately in the morning, the program officers joined the meeting in the afternoon and a most positive and generative discussion ensued.

In Ethiopia, Keystone in partnership with Oxfam America and the Ethiopian Ministry of Agriculture recently surveyed a sample of farmers in six regions on their experience and perception of the government agricultural extension service. We also surveyed the frontline extension officers on their working conditions and the support they receive from the Provincial Agriculture Departments. We recently reported the results to a range of government and civil society stakeholders and are about to return for a round of report-back dialogues with the local Development Agents and with farmers. We are about to begin a three year pilot in South Africa to survey the primary constituents of four South African human rights organizations, and take them through the entire cycle of survey, analysis, report back, dialogue and on-going micro-surveys.

Keystone is on a steep learning curve, but it is our experience so far that organizations that move beyond data collection to dialogue not only learn and improve their performance, they also improve their credibility and achieve higher response rates and more frank feedback in future surveys. One such example comes from an international NGO that has done two rounds of CV surveys with Keystone. While some of their performance scores actually deteriorated over time, their credibility score for listening, responding and engaging with their constituents improved, emphasizing the value constituents placed on the feedback process, and how they appreciated the chance to enter into honest dialogue with the organization.

Constituent Voice in large scale development programs

Aggregating and comparing local performance in provincial or national programs

High quality, 'real-time' performance data for different management levels
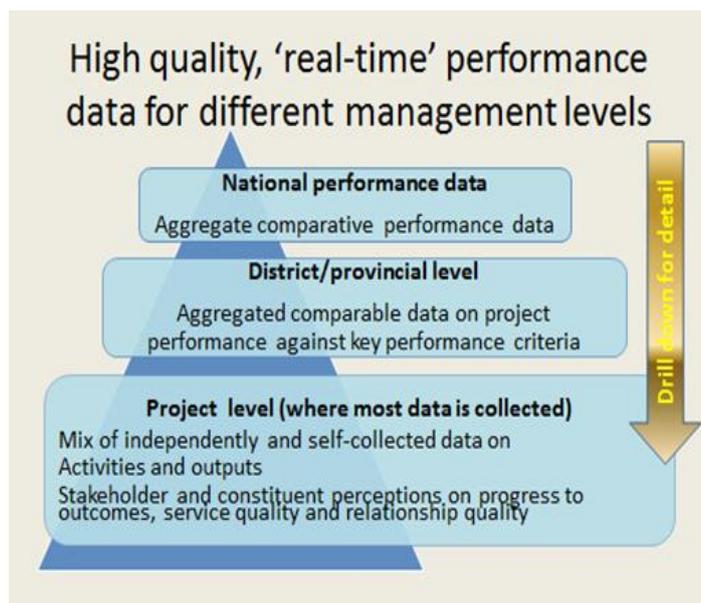
Where there are large scale interventions, quantification enables the aggregation of feedback data into a single database to assess and compare performance of entire programs or departments – right up to national level.

Within the program, CV can amplify the voices of frontline staff by generating feedback on how they are supported by senior management (e.g. how frontline extension staff are supported by their managers, trainers and suppliers of inputs).

For example in Ethiopia, Keystone is partnering with Oxfam America and the Ethiopian Ministry of Agriculture to design a feedback-based

performance management system for the Agricultural Extension Service to make it more farmer-driven, and also to help ensure that local-level Development Agents (DAs) get the support and resources they need from provincial offices to provide quality services to farmers.

In between the comprehensive surveys, a series of micro surveys will generate on-going feedback data from farmers and from local level extension agents on the quality and relevance of services, relationship quality and perceived contributions to outcomes.

Understanding social impact and risk management for companies

Regular CV measurement can help companies better understand and manage the social impacts of their operations and provide early warning signals of problems enabling timely interventions to solve them.

A good illustration of this comes from two sustainable forestry companies in Tanzania and Swaziland that have engaged Keystone to help design and implement a survey-based system for understanding their impact on their surrounding communities. The findings have helped them understand the nature of their relationship with their surrounding communities and making their community development initiatives more responsive to community needs (where previously these were negotiated only with political elites). They have also used the surveys to monitor compliance and employment practices of their local labour contractors and to mitigate risks such as labour disputes and fires started by disgruntled communities.

Correlation of community impacts on company forests with community perceptions of the company
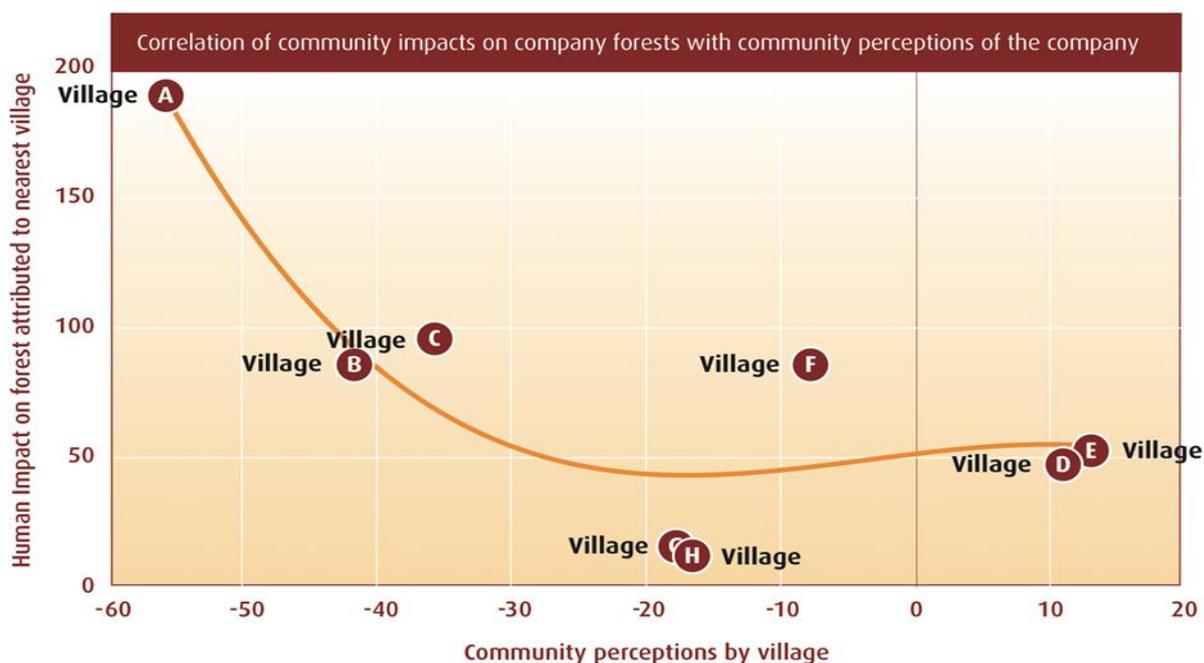
This graph is based on actual data from a Keystone survey for a Tanzanian forestry company. The data showed that poor perceptions of the company were closely related to destructive behaviour such as burning and poaching. The company is now using CV feedback as an early warning signal to guide relationship building efforts.

Similarly, while working for a coffee buyer in Nicaragua, Keystone's CV feedback data enabled the company to provide more of the services that were being truly valued by local farmers, and as such allowed them to increase their market share of total crop sold. The CV methodology gave the company affordable, valuable and easily digestible data, which provided it with clear actionable steps in order to maximise its success.

In conclusion

Like any art or science, Constituent Voice can be done badly. And to unlock its real power will take a great deal more application, reflection and learning by many actors. Over two years Keystone believes that we have basic proof of concept. We have made many mistakes, but also learned a lot. And each time we do it, we and our clients and implementing partners learn more about what works well and the value it generates.

But there are risks. Honest feedback can make us uncomfortable and it can bring simmering conflicts into the open before the solutions become apparent. Feedback can also be manipulated by interest groups. This can make it especially difficult to complete the loops – and most of us need little persuasion to retreat back into our comfort zones.

Another challenge is cost. We know that learning is always the first casualty when organizations barely have the resources they need to carry out their core activities. But the cost of not learning can be much heavier as testified by the empty shells of failed projects littered across the country.

Constituency Voice is no silver bullet, but we are convinced now that the benefits of rigorously collecting and acting on constituent voice are well worth the investment in that they lead to better decisions, better relationships and better outcomes.