FEED BACK LABS PRESENTS

SMART SUMMIT

DISCUSSION PAPER

DRAFT FOR FEEDBACK

This work was made possible by financial and intellectual support from the Fund for Shared Insight.

IS FEEDBACK SMART?

Elina Sarkisova

MAY 2016

FEED BACK LABS

CREDITS

Financial support for the work of Feedback Labs has been provided by:
Fund for Shared Insight, The William and Flora Hewlett Foundation, Bill & Melinda Gates foundation, Rita Allen Foundation, The World Bank, USAID.

Founding Members of Feedback Labs:
Development Gateway, GlobalGiving, Keystone Accountability, Twaweza, Ushahidi, Center for Global Development, Ashoka, FrontlineSMS, GroundTruth.

Collaborating Members:
LIFT, How Matters, Great Nonprofits, CDA, Integrity Action, Results for Development, Charity Navigator, Accountability Lab, Giving Evidence, Global Integrity, Reboot, Voto Mobile, Internews, ICSO.

The author would like to thank Renee Ho and Sarah Hennessy for their helpful input and direction over the course of writing this paper.

DISCLAIMERS

This paper is a draft for discussion and feedback. It does not represent the official views of any referenced organization.

TABLE OF CONTENTS

PREFACE

Aid agencies and philanthropic institutions spend hundreds of billions of dollars each year to improve the well-being of people. Much of this funding is specifically intended to help markets and governments work better. Yet, in recent years, a paradox has emerged: most aid and philanthropy systems are based on relatively closed, top-down, planning-based processes. Few agencies and foundations utilize the type of inclusive, feedback-based design that generates the best results in economic and political spheres.

A group of practitioners, funders, policy makers, researchers, and technologists created Feedback Labs in 2013 as a space to conceptually and operationally address this paradox. Most members and supporters of the Labs instinctively believe that feedback is the "right" thing to do in aid and philanthropy: after all, shouldn't we listen to the people we seek to serve? But this draft paper is a first attempt to explore whether, and under what conditions, feedback is the "smart" thing to do – i.e., whether it improves outcomes in a way that is measurable.

Why do we call this a "draft?" Because we need your feedback. Defining and answering the questions discussed here can't and shouldn't be done by a small group of people. Instead, better understanding will require an open, inclusive, and ongoing conversation where your feedback informs future versions of this paper, additional research, and experimentation.

Let us hear from you at SMART@Feedbacklabs.org

Dennis Whittle
Director, Feedback Labs
April 2016

EXECUTIVE SUMMARY

The idea that the voice of regular people – not only experts – should drive the policies and programs that affect them is not new. Feedback from the people themselves is increasingly seen by aid agencies and non-profits as the right thing to do morally and ethically. But there is much less understanding and consensus about the instrumental value of feedback. Gathering and acting on feedback takes resources, so we must ask if it leads to better outcomes – in other words, is it the smart thing to do? If so, in what contexts and under what circumstances? This paper is a first attempt to frame the issue conceptually, review existing empirical work, and suggest productive avenues for future exploration.

Part of the challenge is to think clearly about what type of feedback is desirable at what stage and for what purpose. Feedback can be collected passively from existing or new data sources (the focus of many "big data" initiatives), or actively from people's perceptions about what they need and the impact of programs on their lives (a question of particular interest to Feedback Labs members). Feedback can also come ex-ante (What programs should we create, and how should we design them?) as well as during implementation (How are things working out? What changes are needed?) and even in evaluation (Looking back, was that a success? What did we learn that will inform what we do differently next time?)

This paper acknowledges, but does not attempt to tackle, all of those issues. It starts by outlining several mechanisms or pathways through which we might expect the incorporation of feedback to lead to better development outcomes:

1. **Knowledge.** Feedback is rooted in important tacit and on-the-ground knowledge essential for a local, contextual understanding. Constituent ownership of a development project would ensure important tacit knowledge makes its way into program design and implementation; but, as donors are the de facto owners, capturing subjective voice offers the next best alternative.
2. **Learning.** Broken feedback loops between donors and constituents, often caused by political and geographical separation, limits an organization's ability – and incentive – to learn. Since knowledge and learning determine the effectiveness of an aid organization, feedback can help organizations learn how to improve their service or intervention.
3. **Adoption.** Getting people to adopt any kind of change or innovation requires their active engagement and participation. The feedback "process" itself can help build trust and lend legitimacy to the intervention, which affects behavior and uptake.

Although case studies and research showing that "x feedback led to y impact" are still few and far between, a handful of studies suggest that feedback can have significant impact on outcomes in some contexts:

- **Reductions in child mortality.** In 9 districts across Uganda, researchers assigned a unique citizen report-card to 50 distinct rural communities. Health facility performance data – based on user experiences, facility internal records and visual checks – was evaluated across a number of key indicators: utilization, quality of services and comparisons to other providers. The report cards were then disseminated during a series of facilitated meetings of both users and providers. Together, they developed a shared vision to improve services and an action plan for the community to implement. As a result, the researchers found significant improvements in both the quantity and quality of services as well as health outcomes, including a 16% increase in the use of health facilities and a 33% reduction in under-five child mortality.
- **Better educational outcomes.** 100 rural schools in Uganda were evaluated, via report card, on improved education-related outcomes. Schools where community members developed their own indicators showed an 8.9% and 13.2% reduction in pupil and teacher absenteeism (respectively) and a commensurate impact on pupil test scores of approximately 0.19 standard deviations (the estimated impact bringing a student from the 50th to the 58th percentile of the normal distribution). In contrast, schools given a report card with standard indicators (developed by experts) showed effects indistinguishable from zero.

While this evidence shows promise, numerous other studies fail to demonstrate that feedback had a measurable impact. One commonly identified problem was that feedback loops did not actually close; people's voices were solicited but not acted on in a way that changed the program. In other cases, even when the feedback loop was closed, factors such as personal bias, access to relevant information, and technical know-how seemed to reduce or negate any possible positive impact. For example:

- Personal bias played a role in citizen satisfaction of water supply duration in India. Satisfaction tended to increase with the hours per day that water was available; however, knowledge of how service compared to that of their peers significantly affected citizens' stated satisfaction. In Bangalore, increasing the number of hours a community had access to water (from one-third of the hours as their neighbors to an equal number of hours) increased the probability of being satisfied by 6% to 18%. However, adding one hour of water access per day increased the probability of being satisfied by only about 1%.
- Technical difficulty affected results in a study on a village-level infrastructure project in Indonesia. Citizens did not believe there was as much corruption in a road-building project in their village as there actually was. Villagers were able to detect marked-up prices but appeared unable to detect inflated quantities of materials, which is where the vast majority of corruption in the project occurred. Grassroots "bottom-up" participation in the monitoring process yielded little overall impact, but introducing a "top-down" government audit reduced the missing expenditures by 8%.

One finding of the paper is that the feedback process itself, when done well, can help build trust and legitimacy, often a necessary condition for people to adopt even interventions that are designed top-down by experts. The research reviewed also suggests that people are often not only good judges of the quality of services they receive, but that their subjective assessments are based on a more inclusive set of factors than is otherwise possible to capture with standard metrics.

The bottom line seems to be that feedback can be smart when people are sufficiently empowered to fully participate, when the technical conditions are appropriate, and when the donor and/or government agency has both the willingness and capacity to respond. But much work remains to be done to flesh out the exact conditions under which it makes sense to develop and deploy feedback loops at different stages within different types of programs.

This report suggests several principles to guide future exploration of this topic: (1) Use a variety of different research approaches – from randomized control trials (RCTs) to case studies – to further build the evidence base; (2) Explore different incentives and mechanisms – both on the supply and demand sides – for "closing the loop"; (3) Test different ways of minimizing bias and better understanding the nature of information that empowers people; and (4) Conduct more cost-benefit analysis to see whether feedback is not only smart but also cost-effective at scale. But the most important principle is 5) Seek feedback from the community itself – you, the reader – about the paper's initial findings and ideas for future research and experimentation.

You can send us that feedback at SMART@FeedbackLabs.org.
We promise to close the loop.

INTRODUCTION

This paper is motivated by the idea that regular people – not experts – should ultimately drive the policies and programs that affect them. The idea is not new. In fact, it underpins a number of important efforts (both old and new) among development partners to "put people first," including user satisfaction surveys, community engagement in resource management and service delivery (participatory and community development), empowering citizens vis-a-vis the state (social accountability) and enabling greater flexibility, learning and local experimentation in the so-called "science of delivery" through new tools like problem-driven iterative adaptation (PDIA). It is this basic idea that drives the work of Feedback Labs.

This paper asks if feedback is smart, or results in better social and/or economic outcomes for poor people.

However, while the idea that people's voices matter is not new, few organizations systematically collect – and act on – feedback from their constituents and even fewer know how to do it well.[1] We think this stems in part from a lack of clarity around its instrumental value. In other words, aside from it being the ***right*** thing to do (something most people can agree on), is it also the ***smart*** thing? Given the mixed evidence on many feedback-related initiatives,[2] [3] coupled with the reality that aid dollars are a finite resource that now more than ever needs to be guided by good evidence, it seems like a reasonable question to ask. This report is our attempt to shed light on this question, not to offer conclusive answers but rather to spark an ongoing conversation and inform future experimentation and research.

The rest of the report is split into four parts. The first reviews some key theoretical literature to help explain why we might expect feedback to be the smart thing. What are some of the key mechanisms or pathways through which the incorporation of constituent voice (or the "closed loop") might lead to better development outcomes (i.e. improved student learning, reductions in childhood mortality, etc.)? The second explores whether there is any evidence to suggest that it actually does. The third attempts to make sense of evidence that suggests it does not. We rely on experimental (i.e. randomized control trials (RCTs)) or quasi-experimental approaches where we can but use case studies and other approaches to fill in gaps. The review is not meant to be exhaustive but rather highlight some general ideas and themes. In the concluding section, we summarize our findings and suggest areas for further research.

1. At Feedback Labs, we recognize that simply collecting feedback is not enough. Constituents must be actively engaged in every step of the project cycle – from conception to evaluation – and their feedback must be used to influence decision-making. Moreover, who participates matters: all relevant stakeholders must be successfully brought in. For a more detailed description of the steps required in a closed feedback loop please visit our website: www.feedbacklabs.org.

But first, what exactly do we mean by feedback?

We focus on subjective or "perceptual" feedback that is voiced directly from constituents.

There are a variety of forms of feedback that can be useful in improving outcomes. The collection, analysis, and use of "big" data have exploded in recent years. Much of this data is collected passively, often using new cost-effective digital tools. In this paper, we focus on another – less studied – form of feedback that is:

- **Voiced directly from regular individuals who are the ultimate intended recipients of social and economic programs (hereafter referred to as "constituents").** We exclude feedback from policy makers or government officials; while they are sometimes the intended recipients of aid (either directly, as the case of technical assistance programs, or indirectly, when external funding has to pass through or be managed by a government agency), our main focus is on the people whose well-being the aid is ultimately intended to improve.

- **Subjective, or "perceptual," in nature (i.e. speaking to the person's opinions, values or feelings).** Examples of perceptual feedback on a service might include "I benefited a lot from this service" or "This service was good." This is distinct from feedback that provides more objective information or data that can simply be collected from a person rather than actively voiced. An example might include a software application that tracks a person's behavior (i.e. the number of footsteps in a day) and relays it to the person or her physician.

- **Collected at any stage of a program**, including conception, design, implementation or evaluation.

- **Deliberately collected or "procured."** While there may be value in unsolicited or spontaneous feedback, our paper focuses on feedback that is collected deliberately.

We make two important assumptions. The first – which we return to in greater detail in section three – is that feedback loops actually close. For that to happen, many different pieces need to fall into place: on the demand-side, the right people must actually participate, or offer their feedback, in the first place. We know this is not always the case, as participation has economic – and often political – costs which disadvantage some more than others, an issue often ignored in many donor programs. Participation also suffers from free rider problems, as benefits are non-excludable. Still more, feedback must be adequately aggregated and represented and finally translated into concrete policy outcomes, which may be particularly challenging in highly heterogeneous societies where people disagree about the best course of action.

In asking if feedback is smart, we assume that feedback loops actually close...

On the supply side, the entity (donor or government) on the receiving end of the feedback must be both willing and able to act on it. We recognize that despite a proliferation of efforts (among both donors and governments) to more actively engage citizens, this too is not always the case. For these reasons, a number of authors find that the vast majority of donor feedback-related initiatives fail to achieve their intended impact. The purpose of this paper is not to try to explain why feedback loops do, or do not, close but rather that when they do, we see improved outcomes. However, to ignore these issues completely would be to look at the issue with one eye shut. For this reason, we briefly return to some of these issues in the third section.

...and that the "experts" know what regular people need to make their lives better. We know that neither of these assumptions is always true.

The second is that we acknowledge a deep-rooted assumption that experts know what people need to make their lives better. The measured outcomes of interest – i.e. the desired social or economic outcomes – are usually identified and specified in advance and in a top-down fashion by experts who may not know what constituents actually want. We note the problems in this assumption and disagree that "experts always know best." This is an issue we will explore in a future paper. For the scope of this paper, we operate within the existing world of aid and philanthropy and accept the specified outcomes of interest as experts specify them.

4. For a more thorough analysis of this please refer to Mansuri and Rao (2013); Gaventa and McGee (2013); Fox (2014); and Peixoto and Fox (2016).

Section One

WHAT DOES THE THEORY SAY?

In this section, we ask the question, "Why we would we ***expect*** feedback to be the smart thing?" What are the mechanisms or pathways through which the incorporation of constituent voice leads to better development outcomes? A broad review of the literature points to three potential pathways: (1) tacit or time-and-place knowledge, (2) organizational learning, and (3) legitimacy. First, feedback is rooted in tacit or time-and-place knowledge that is essential for understanding the local context. Second, feedback can help organizations learn how to improve their service or intervention. Third, the feedback "process" itself can help build trust and lend legitimacy to the intervention, which affects behavior and uptake of the intervention.

Feedback is rooted in important tacit or time-and-place knowledge that is essential for understanding the local context.

Most development practitioners recognize that solving complex development challenges requires a deep understanding of local conditions and context. However, doing so requires tapping into intangible forms of knowledge that are difficult to transfer using traditional tools of scientific inquiry. The economist and philosopher Friedrich Hayek referred to this as "time-and-place knowledge"[5] and argued that it cannot, by its nature, be reduced to simple rules and statistical aggregates and, thus, cannot be conveyed to central authorities to plan an entire economic system.[6] Instead, it stands the best chance of being used when the individuals in possession of this knowledge are themselves acting upon it or are actively engaged.

Feedback offers the best chance we have for ensuring that important tacit and time-and-place knowledge gets incorporated into program design and implementation.

Similarly, scientist and philosopher Michael Polanyi recognized the existence of tacit knowledge, knowledge that is conceptual and/or sensory in nature and is difficult to transfer using formal methods. It comprises informed guesses, hunches, and imaginings. As he wrote in his 1967 classic *The Tacit Dimension*, "we can know more than we can tell." Both tacit and time-and-place knowledge contrast with scientific – or technical – knowledge that can be written down in a book.

5. We refer to Hayek's "time-and-place" theory similarly, as well as with the phrase "on-the-ground"
6. Hayek (1945)

It is not to say that one type of knowledge trumps the other. In development – indeed in any human endeavor – one needs a combination of both.[7] Several authors argue that the importance of soft or intangible knowledge increases under conditions of uncertainty, or where there is a weak link between inputs and desired outcomes, a condition that applies to most development problems.[8] However, most policymakers – including donors – place greater value on scientific knowledge over and above tacit or time-and-place knowledge even when conditions might warrant a greater focus on the latter.

Political economist Elinor Ostrom argues that, given its intangible nature, giving locals "ownership" over development programs offers the best chance that tacit or time-and-place knowledge will be incorporated into the design and implementation of projects.[9] Although she acknowledges that the theoretical conception of ownership is not entirely clear she argues that giving constituents ownership involves allowing them to articulate their own preferences and be more actively engaged in the provision and production of aid, including shared decision-making over the long-term continuation or non-continuation of a project.[10] However, given that donors will always be the *de facto* owners by virtue of the fact that they pay for programs, true constituent ownership remains an elusive ideal. By capturing subjective voice, feedback offers the next best alternative for ensuring that important tacit and time-and-place knowledge make their way into program design and implementation.

Contribute www.smartsummit_url.com

7. Ibid. Andrews,
8.Pritchett and Woolcock (2012).
9. Ostrom (2001: 242-243)
10. Ostrom identifies four dimensions of ownership: (1) enunciating demand, (2) making a tangible contribution, (3) obtaining benefits, and (4) sharing responsibility for long-term continuation or non-continuation of a project. (Ostrom (2001:15)).

Feedback can help organizations learn about how to improve their service or intervention.

In *Aid on the Edge of Chaos*, Ben Ramalingam argues that the effectiveness of organizations is central to social and economic development and that knowledge and learning are the primary basis of their effectiveness.[11] Yet most aid organizations do not know how to learn. Chris Argyris, one of the pioneers of organizational learning, distinguishes between two types of learning: single-loop learning, or learning that reinforces and improves existing practices, and double-loop learning, or learning that helps us challenge and innovate.[12] To illustrate single-loop learning, he uses the analogy of a thermostat that automatically turns on the heat whenever the temperature in a room drops below a certain pre-set temperature.

The political and geographical separation between donors and constituents gives rise to a broken feedback loop, which seriously limits aid agencies' ability – and incentives – to learn. Repairing the loop is central to achieving outcomes.

In contrast, double-loop learning requires constant questioning of existing practices by rigorously collecting – and responding to – feedback. To return to the thermostat example, double-loop learning would require questioning the rationale for using the pre-set temperature and adjusting it accordingly. Noted systems thinker Peter Senge gives the example of learning to walk or ride a bike: we learn through multiple attempts and bodily feedback – each time our bodies, our muscles, our sense of balance react to this feedback by adjusting to succeed.[13] This is also the basic idea behind "problem-driven iterative adaptation" (PDIA), a new approach to helping aid agencies achieve positive development impact.[14] The authors argue that, given the complexity of solving development challenges, outcomes can only "emerge" as a puzzle gradually over time, the accumulation of many individual pieces. Thus, solutions must always be experimented with through a series of small, incremental steps involving positive deviations from extant realities, a process Charles Lindblom famously calls the "science of muddling through."[15] This kind of experimentation has the greatest impact when connected with learning mechanisms and iterative feedback loops.

However, despite the centrality of active, iterative learning in development, few aid agencies actually do it. Ramalingam points to two main reasons: cognitive and political. On the one hand, organizations are inhibited by what Argyris calls "defensive reasoning," or the natural tendency among people and organizations to deflect information that

11. Ramalingam (2013: 19)
12. Argyris (1991).
13. Senge (1990).
14. Andrews, Pritchett and Woolcock (2012).
15. Lindblom (1959).

puts them in a vulnerable position. The other is political – namely, conventional wisdom becomes embedded in particular institutional structures which then act as a "filter" for real knowledge and learning. In other words, "power determines whose knowledge counts, what knowledge counts and how it counts" and becomes self-perpetuating.[16]

In the private sector, companies that do not engage in active feedback and learning may lose customers and eventually go out of business. This is because the customer pays the company directly, can observe whether the product or service meets his/her expectations and – if dissatisfied – has the power to withhold future business. In contrast, in aid and philanthropy, the people who are on the receiving end of products and services – the constituents – are not the same people who actually pay for them – i.e. taxpayers in donor countries. Moreover, donors and constituents are often separated by thousands of miles, making information about how programs are progressing difficult – if not impossible – to obtain. This political and geographical separation between donors and constituents gives rise to a broken feedback loop, which seriously limits aid agencies' ability – and incentives – to learn. Owen Barder argues that instead of fighting this, development partners should just accept it as an inherent characteristic of the aid relationship and focus their energy on building collaborative mechanisms and platforms that help repair the broken feedback loop.[17]

Getting people to adopt any kind of change or innovation requires their active engagement and participation. Feedback can help facilitate this process and build legitimacy.

Contribute

www.smartsummit_url.com

16. Ramalingam (2013: 27).
17. Barder (2009).

Feedback can help build trust and lend legitimacy to the intervention, which is key for successful implementation.

It is one thing to come up with an innovative product or service – it is quite another to get its intended recipients to actually adopt it. While much traditional social theory is built on the assumption that behavior is motivated by rewards and punishments in the external environment, legitimacy has come to be regarded as a far more stable – and not to mention cost-effective – base on which to rest compliance. For instance, a host of social scientists argue that citizens who accept the legitimacy of the legal system and its officials will comply with their rules even when such rules conflict with their own self-interest.[18] In this way, legitimacy confers discretionary authority that legal authorities require to govern effectively. However, this is not unique to the law. All leaders need discretionary authority to function effectively, from company managers who must direct and redirect those who work under them to teachers who want their students to turn in their homework assignments on time. Donors are no exception.

What builds legitimacy? Borrowing from the literature on institutional change, Andrews, Pritchett and Woolcock highlight the importance of broad participation in ensuring legitimacy.[19] They argue that getting people to adopt any kind of change requires "the participation of all actors expected to enact the innovation" and especially "the more mundane and less prominent, but nonetheless essential, activities of 'others,'" also referred to in the literature as "distributed agents."[20] These 'others' need to be considered because if institutionalized rules of the game have a prior and shared influence on these agents – i.e. if they are institutionally "embedded" – they cannot be expected to change simply because some leaders tell them to. Instead, they must be actively engaged in two-way "dialogue that reconstructs the innovation as congruent with [their] interests, identity and local conditions."[21] Feedback can help facilitate this kind of dialogue.

18. Tyler (1990).
19. Andrews, Pritchett and Woolcock (2012).
20. Whittle, Suhomlinova and Mueller (2011: 2)
21. Ibid.

Contribute

www.smartsummit_url.com

Section Two

WHAT DOES THE EVIDENCE SAY ABOUT WHETHER FEEDBACK IS THE SMART THING?

In the previous section, we explain why we would – in theory – expect feedback to result in better development outcomes. In this section, we explore the evidence for whether it actually does. In other words, does incorporating feedback from constituents – whether they are students, patients or other end users of social services – actually result in better outcomes (i.e. improved student learning, reductions in childhood mortality, etc.)? We are interested primarily in studies that construct a counterfactual scenario using experimental (randomized control trials (RCTs)) or quasi-experimental methods. We do, however, rely on some qualitative assessments to fill in gaps. This review is not meant to be exhaustive but rather to highlight some general themes and ideas that we hope will inform future research and experimentation.

The evidence for feedback has not yet caught up to theory and practice but it is beginning to emerge.

Direct evidence of impact

In the development context, perhaps some of the strongest evidence exists in the area of community-based monitoring. To test the efficacy of community monitoring in improving health service delivery, researchers in one study randomly assigned a citizen's report card to half of 50 rural communities across 9 districts in Uganda.[22] The report card, which was unique to each treatment facility, ranked facilities across a number of key indicators, including utilization, quality of services and comparisons vis-à-vis other providers. Health facility performance data was based on user experiences (collected via household surveys) as well as health facility internal records[23] and visual checks. The report cards were then disseminated during a series of facilitated meetings of both users and providers aimed at helping them develop a shared vision of how to improve services and an action plan or contract – i.e. what needs to be done, when, by whom – that was then up to the community to implement.

22. Bjorkman and Svensson (2007)
23. Because agents in the service delivery system may have a strong incentive to misreport key data, the data were obtained directly from the records kept by facilities for their own need (i.e. daily patient registers, stock cards, etc.) rather than from administrative records.

A citizens' report card in Uganda led to a 16% increase in utilization and a 33% reduction in under-five child mortality.

The authors of the study found large and significant improvements in both the quantity and quality of services as well as health outcomes, including a 16% increase in the use of health facilities and a 33% reduction in under-five child mortality. However, while these results are promising, it is difficult to know precisely through which channel feedback actually worked. For instance, it could have been the act of aggregating and publicly sharing user feedback on the status of service delivery, which could have helped the community manage expectations about what is reasonable to expect from providers. It could also have been through the participatory mechanism itself – i.e. mobilizing a broad spectrum of the community to contribute to the performance of service providers.

In part to help close this gap, researchers in another study evaluated the impact of two variations of a community scorecard – a standard one and a participatory one – to see which of them led to improved education-related outcomes among 100 rural schools in Uganda.[24] In schools allocated to the standard scorecard, scorecard committee members (representing teachers, parents, and school management) were provided with a set of standard indicators developed by experts and asked to register their satisfaction on a 5-point scale. They were then responsible for monitoring progress throughout the course of the term. In contrast, in schools allocated to the participatory scorecard, committee members were led in the development of *their own* indicators to rate and monitor. This participatory aspect of the scorecard was the only difference between the two treatment arms.

In another experiment, a report card initiative that allowed constituents to design their own indicators outperformed the standard one.

Results show positive and significant effects of the participatory design scorecard across a range of outcomes: 8.9% and 13.2% reduction in pupil and teacher absenteeism (respectively) and a commensurate impact on pupil test scores of approximately 0.19 standard deviations (the estimated impact of approximately 0.2 standard deviations would raise the median pupil 8 percentage points, or from the 50th to the 58th percentile of the normal distribution). In contrast, the effects of the standard scorecard were indistinguishable from zero. When comparing the qualitative choices of the two scorecards, researchers found that the participatory scorecard led to a more *constructive*

framing of the problem. For example, while there was broad recognition that teacher absenteeism was a serious issue (one that gets a lot of focus in the standard economics literature), the participatory scorecard focused instead on addressing its root causes – namely, the issue of staff housing in rural areas (requiring teachers to travel long distances to get to school and be more likely absent). Thus, researchers attribute the relative success of the participatory scorecard to its success in coordinating the efforts of school stakeholders – both parents and teachers – to overcome such obstacles.

Moving beyond the strictly development context, there is movement in psychotherapy towards "feedback-informed treatment," or the practice of providing therapists with real-time feedback on patient progress throughout the entire course of treatment...*but from the patient's perspective.*[25] It turns out that asking patients to subjectively assess their own well-being and incorporating this information into their treatment results in fewer treatment failures and better allocative efficiency (i.e. more at-risk patients end up getting more hours of treatment while less at-risk patients get less).[26] Moreover, providing therapists with additional feedback – including the client's assessment of the therapeutic alliance/relationship, readiness for change and strength of existing (extra-therapeutic) support network – increases the effect, doubling the number of clients who experience a clinically meaningful outcome.[27]

In psychotherapy, asking patients to subjectively assess their own wellbeing and incorporating this feedback into their treatment results in fewer treatment failures and better allocative efficiency.

In another study, patient-centered care, or "care that is respectful of and responsive to individual patient preferences, needs, and values and ensures that patient values guide all clinical decisions,"[28] was associated with improved patients' health status and improved efficiency of care (reduced diagnostic tests and referrals).[29] This relationship was both statistically and clinically significant: recovery improved by 6 points on a 100-point scale and diagnostic tests and referrals fell by half. However, only one of the two measures used to measure patient-centered care was linked to improved outcomes: the patients' perceptions of the patient-centeredness of the visit. The

25. Minami and Brown.
26. At least five large RCTs have been conducted evaluating the impact of patient feedback on treatment outcomes. These findings are consistent across studies. See Lambert (2010) for review.
27. Lambert (2010: 245).
28. Defined by the Institute of Medicine (IOM), a division of the National Academies of Sciences, Engineering, and Medicine. The Academies are private, nonprofit institutions that provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions related to science, technology, and medicine. The Academies operate under an 1863 congressional charter to the National Academy of Sciences, signed by President Lincoln. See more at: http://iom.nationalacademies.org
29. Stewart et al (2000)

other (arguably more objective) metric – the ratings of audiotaped physician-patient interactions by an independent third party – was not directly related, suggesting that patients are able to pick up on important aspects of care that are not directly observable but consequential to the outcomes of interest (something we build on below).

According to one study, medical treatment that took into account patient preferences, needs and values resulted in a 6% improvement in recovery and 50% reduction in referrals.

Contribute

www.smartsummit_url.com

Indirect evidence of impact

In addition, a number of studies provide indirect evidence that feedback is the smart thing, by corroborating some of the pathways identified in the theoretical review. First, the predictive power of self-rated health suggests that people are unusually good – better than experts give them credit for – at assessing their own problems. Second, the evidence for user satisfaction surveys suggests that people are also – by extension – good judges of the quality of services being delivered by experts. Last, a number of studies suggest that, when properly implemented, the feedback process itself – of conferring, listening, bringing people in, etc. – can build trust, which can lead to positive behavior change, thus contributing to improved outcomes.

Self-rated health (SRH) – also known as self-assessed health, self-evaluated health, subjective health or perceived health – is typically based on a person's response to a simple question: "How in general would you rate your health – poor, fair, good, very good or excellent?" A number of studies have shown that SRH – contrary to the intuition that self-reporting diminishes accuracy – is actually a strong predictor of mortality and other health outcomes. Moreover, in most of these studies, SRH retained an independent effect even after controlling for a wide range of health-related measures, including medical, physical, cognitive, emotional and social status.[30] Experts argue that its predictive strength stems precisely from its subjective quality (the very quality skeptics criticize it for) – namely, when asked to rate their own health individuals consider a more inclusive set of factors than is usually possible to include in a survey instrument or even to gather in a routine clinical examination.[31] This suggests that subjective metrics could be particularly well suited for measuring outcomes that are multi-dimensional.

Self-rated health is predictive of health outcomes. Its strength comes from its subjective quality – namely, when asked to rate their own health, individuals consider a more inclusive set of factors than is otherwise possible with routine clinical procedures.

While it makes sense on some level why such an internal view would be privileged with respect to one's *own* health, are people also good judges of the performance of highly trained professionals, whether they are doctors, teachers or government bureaucrats? This is important because if, as we suggest in our theoretical review, feedback can help organizations improve their services, performance should be the main driver

30. See Schnittker and Bacak (2014) for review.
31. Benyamini (2011)

of perceived service delivery and satisfaction. A review of the evidence around user satisfaction surveys suggests that people are not only good judges of the value of services being delivered, they are also able to pick up on important aspects of care that are otherwise difficult to measure.

Patient satisfaction with care is positively correlated with mortality. Patients' qualitative assessments are sensitive to factors not well captured by clinical performance metrics.

According to one study, controlling for a hospital's clinical performance, higher hospital-level patient satisfaction scores were associated with lower hospital inpatient mortality rates, suggesting that patients' subjective assessment of their care provides important and valid information about the overall quality of hospital care that goes beyond more objective clinical process measures (i.e. adherence to clinical guidelines).[32] Specifically, it found that (1) the types of experiences that patients were using when responding to the overall satisfaction score were more related to factors one would normally expect to influence health outcomes (i.e. how well doctors kept them informed) rather than superficial factors like room décor; and (2) patients' qualitative assessments of care were sensitive to factors not well captured by current clinical performance metrics but that have been linked with patient safety and outcomes – for instance, the quality of nursing care.

Still more, studies show that well-designed student evaluations of teachers (SETs) can provide reliable feedback on aspects of teaching practice that are predictive of student learning. The Measures of Teaching project surveyed the perceptions of 4th to 8th graders using a tool that measured specific aspects of teaching.[33] They found that some student perceptual data was positively correlated with student achievement data – even more so than classroom observations by independent third parties. Most important are students' perception of a teacher's ability to control a classroom and to challenge students with rigorous work – two important areas of teacher effectiveness that arguably only students can truly judge.

One study involving U.S. middle school students showed a positive correlation between student perceptions of teacher effectiveness and their performance on exams.

32. Glickman et al (2010)
33. MET project (2013).

Last, a number of studies in the area of natural resource management suggest that – when properly implemented – feedback can indeed build trust, which can then lead to positive behavior change, contributing to improved outcomes. In one study, both quantitative and qualitative methods were used to assess the impact of two participatory mechanisms – local management committees and co-administration – on the effectiveness of Bolivia's Protected Areas Service (SERNAP) in managing its protected areas, as measured by five broad areas: overall effectiveness, basic protection, long-term management, long-term financing and participation.[34] It found that both participatory mechanisms helped SERNAP improve protected areas management (as compared with areas that that it managed on its own), not only by enabling authorities to adapt instructions to local context but also by building trust.

Trust has been shown to be one of the most consistent predictors of how local people respond to a government-protected area. A linear regression analysis of 420 interviews with local residents living within the immediate vicinities of three major U.S. national parks revealed that process-related variables (i.e. trust assessments, personal relationships with park workers and perceptions of receptiveness to local input) overpowered purely rational assessments (i.e. analysis of the costs vs. benefits associated with breaking park rules) in predicting the degree to which respondents actively supported or actively opposed each national park.[35] [36] The magnitude was large (explaining 55-92% of the variation in local responses) and consistent across all three parks. Moreover, perceptions of the trustworthiness of park managers were the most consistent explanatory variable in the study, especially in explaining why locals opposed the park.

Participatory approaches improved the effectiveness of Bolivia's Protected Areas Service (SERNAP) in managing its protected areas, primarily by building trust and legitimacy.

34. Mason et al. (2010).
35. Stern (2010).
36. "These actions were measured not only through self-reporting but also through triangulation techniques using multiple key informants and field observation. Park managers were used as Thurstone judges to create a gradient of active responses from major to minor opposition and support. Instances of major active opposition included intentional resource damage or illegal havesting, harassing park guards, filing lawsuits, public campaigning, and/or active protesting against the parks. Major supporting actions included giving donations, volunteering, changing behaviors to comply with park regulations, and/or defending the parks in a public forum. Other actions, however, were categorized as minor support or opposition. For example, dropping just a few coins once or twice into a donation box, picking a few flowers on a recreational visit, or talking against the park informally with friends or coworkers fell into these categories." (Taken directly from Stern (2010)).

While it is difficult to generalize about how trust is built or lost, the researchers performed quantitative analyses on three categories of data in order to reveal the variables most powerfully related to trust at each park: (1) locals' own self-reported reasons for trusting or distrusting park managers, (2) open-ended complaints lodged by respondents against the parks at any point during their interviews, and (3) demographic/contextual variables (i.e. gender, income, education, etc.). Based on conceptualizations of trust in the literature, researchers then categorized trust into two types: rational ("*Entity* A trusts *Entity* B because *Entity* A expects to benefit from the relationship") and social ("*Entity* A trusts *Entity* B because of some common ground or understanding based on shared characteristics, values or experience or upon open and respectful communication"). The study found that although rational assessments were correlated to trust assessments at each park, they were overpowered in each setting by themes related to cultural understanding, respect and open and clear communication. Perceived receptiveness to local input alone explained 20-31% of overall variations in trust in two of the three parks.

Contribute

www.smartsummit_url.com

Section Three

CAVEATS: WHEN IS FEEDBACK NOT THE SMART THING?

While the evidence reviewed above shows promise, numerous other studies suggest that feedback is not always the smart or effective thing. In their review of over 500 studies on donor-led participatory development and decentralization initiatives, Mansuri and Rao show that the majority failed to achieve their development objectives. Similarly, a review of the top 75 studies in the field of transparency and accountability initiatives shows that the evidence on impact is mixed.[37] Jonathan Fox reaches a similar conclusion in his review of 25 quantitative evaluations of social accountability initiatives. In this section, we attempt to shed light on some of the reasons. As noted in previous sections, in most cases, the main reason was because feedback loops did not actually close. On the supply side, there was never the willingness or capacity to respond to feedback in the first place. On the demand side, not everyone participated and/or there were breakdowns in aggregating, and then translating, people's preferences into concrete policy decisions. However, even in cases where all of these conditions are met, other demand-side factors sometimes get in the way – namely, personal bias, access to relevant information and technical know-how (or lack thereof).

Feedback loops do not always close, thus failing to achieve outcomes. And even when they do, bias and other demand-side factors can get in the way.

The state and/or donor must be willing and able to respond

First, not all studies are pursuing the same theory of change. For instance, social accountability is an evolving umbrella term used to describe a wide range of initiatives that seek to strengthen citizen voice vis-à-vis the state – from citizen monitoring and oversight of public and/or private sector providers to citizen participation in actual resource allocation decision making – each with its own aims, claims and assumptions. By unpacking the evidence, Jonathan Fox shows that they are actually testing two very different approaches: tactical and strategic.

37. McGee and Gaventa (2010)

Feedback is smart only when the donor and/or government agency has both the willingness and capacity to respond…

Tactical approaches – which categorize the vast majority of donor-led initiatives – are exclusively demand-side efforts to project voice and are based on the unrealistic assumption that information alone will motivate collective action, which will, in turn, generate sufficient power to influence public sector performance. According to Fox, such interventions test extremely weak versions of social accountability and – not surprisingly – often fail to achieve their intended impact. In contrast, strategic approaches employ multiple tactics – both on the demand and supply sides – that encourage enabling environments for collective action (i.e. independent media, freedom of association, rule of law, etc.) and coordinate citizen voice initiatives with governmental reforms that bolster public sector responsiveness (i.e. "teeth"). Such mutually reinforcing "sandwich" strategies, Fox argues, are much more promising.

Mansuri and Rao reach a similar conclusion in their review of over 500 studies on donor-led participatory development and decentralization initiatives: "Local participation tends to work well when it has teeth and when projects are based on well-thought-out and tested designs, facilitated by a responsive center, adequately and sustainably funded and conditioned by a culture of learning by doing." In their review, they find that "even in projects with high levels of participation, 'local knowledge' was often a construct of the planning context and concealed the underlying politics of knowledge production and use." In other words, while feedback was collected, the intention to actually use it to inform program design and implementation was never really there, resulting in a closed loop.

Contribute

www.smartsummit_url.com

Who participates matters

On the demand-side, the issue of who participates matters. Mansuri and Rao find that participation benefits those who participate, which tend to be wealthier, more educated, of higher social status, male and the most connected to wealthy and powerful people.[40] In part, this reflects the higher opportunity cost of participation for the poor. However, power dynamics play a decisive role. In other words, particularly in highly unequal societies, power and authority are usually concentrated in the hands of a few, who – if given the opportunity – will allocate resources in a way that gives them a leg up. Thus, "capture" tends to be greater in communities that are remote, have low literacy, are poor, or have significant caste, race or gender disparities. Mansuri and Rao find little evidence of any self-correcting mechanism through which community engagement counteracts the potential capture of development resources. Instead, they find that it results in a more equitable distribution of resources only where the institutions and mechanisms to ensure local accountability are robust (echoing the need for a strong, responsive center in leveling the playing field, as discussed above).

. . .and when people are sufficiently empowered to fully participate.

In addition, participation suffers from free rider problems. In discussing the potential of democratic institutions in enhancing a country's productivity, Elinor Ostrom argues that devising new rules (through participation) is a "second-order public good."[41] In other words, "the use of a rule by one person does not subtract from the availability of the rule for others and all actors in the situation benefit in future rounds regardless of whether they spent any time and effort trying to devise new rules." Thus, one cannot automatically presume that people will participate just because such participation promises to improve *joint* benefits. Ostrom argues that people who have interacted with one another over a long time period and expect to continue these interactions far into the future are more likely to do so than people who have not.[42] However, such claims – while compelling in theory – are not always backed by rigorous evidence.

40. Mansuri and Rao (2013: 123-147).
41. Ostrom (2001: 52)
42. Ostrom (2001: 53)

Aggregation and representation

Even if everyone participates, translating preferences into outcomes can be difficult. For instance, while voting is an important mechanism for aggregating preferences – particularly in democratic societies – it is besieged by difficulty. Citing Kenneth Arrow's Impossibility Theorem (1951), Elinor Ostrom argues, "no voting rule could take individual preferences, which were themselves well-ordered and transitive, in a transitive manner and guarantee to produce a similar well-ordered transitive outcome for the group as a whole."[43] When members of a community have very similar views regarding preference orderings, this is not a big problem. However, in highly diverse societies, there is no single rule that will guarantee a mutually beneficial outcome. Thus, "we cannot simply rely on a mechanism like majority vote to ensure that stable efficient rules are selected at a collective-choice level."[44]

The issue of aggregation and representation has been highlighted in the social accountability sphere as well, where much of the emphasis is on aggregating citizen voice through mechanisms like satisfaction surveys or ICT platforms rather than translating it into effective representation and eventually to desired outcomes. According to Fox, "this process involves not only large numbers of people speaking at once, but the consolidation of organizations that can effectively scale up deliberation and representation as well – most notably, internally democratic mass organizations."[45] However, given some of the challenges noted above, which types of decision rules produce the best joint outcomes – particularly in highly heterogeneous societies – is an open one.

Contribute www.smartsummit_url.com

43. Ostrom (2001: 56)
44. Ibid
45. Fox (2014: 26)

Role of personal biases

A number of studies show that bias is not only real but also sometimes difficult to predict and control for. These studies suggest that relying on constituent feedback alone is unlikely to maximize desired outcomes – and in some cases may actually have a negative impact.

One study involving university-level students showed a near-zero correlation between perception data and performance on final exams.

A recent study (comprising of both a natural experiment and randomized control trial) of over 23,000 university-level SETs in France and the U.S. found that they were more predictive of students' grade expectations and instructors' gender than learning outcomes, as measured by performance on anonymously graded, uniform final exams.[46] It found a near-zero correlation between SET scores and performance on final exams (for some subjects it was actually negative but statistically insignificant); in contrast, biases were (1) large and statistically significant and affected how students rated even putatively objective aspects of teaching, such as how promptly assignments are graded; (2) skewed against female instructors; and (3) very difficult to predict and therefore control for (i.e. the French university data show a positive male student bias for male instructors while the U.S. setting suggests a positive female student bias for male instructors). These findings suggest that relying on SETs alone is not likely to be a good predictor of student learning outcomes (and therefore teacher effectiveness).

When looking at optimal weights for a composite measure of teaching effectiveness that included teachers' classroom observation results, SETs, and student achievement gains on state tests, the same Measures of Teaching project mentioned above found that reducing the weights on students' state test gains and increasing the weights on SETs and classroom observations resulted in better predictions of (1) student performance on supplemental (or "higher order") assessments[47] and (2) the reliability[48] of student outcomes…*but only up to a point*. It turns out there's a sweet spot: going up from a 25-25-50 distribution (with 50% assigned to state test gains) to a 33-33-33 (equal) distribution actually reduced the composite score's predictive power.[49] The study supports the finding above that SETs should not be used as the sole source of evaluating teacher effectiveness but when well-designed (i.e. targeting specific aspects of teaching) and used in combination with more objective measures could be more predictive of student learning than using more objective measures alone.

46. Boring and Stark (2016).
47. The MET study measured student achievement in two ways: (1) existing state tests and (2) three supplemental assessments designed to assess higher-order conceptual understanding. While the supplemental tests covered less material than the state tests, the supplemental tests included more cognitively challenging items that required writing, analysis and application of concepts.
48. This refers to the consistency of results from year to year.
49. MET project (2013).

A study involving U.S. middle school students showed that combining perceptual data with more objective metrics resulted in better predictions of student performance than either of those things alone.

Yet another study investigates citizen satisfaction with service delivery – in this case water supply – using household survey data from two Indian cities (Bangalore and Jaipur).[50] The surveys collected detailed information on households' satisfaction with various aspects of water service provision as well as information on actual services provided (i.e. the quantity provided, the frequency at which the service is available and the quality of the product delivered). However, to isolate the impact of household and community characteristics on a household's satisfaction with service provision, the empirical analysis focused only on households' satisfaction with the *duration* of water supply (hours per day), a more objective indicators of service quality.

Citizen satisfaction with the duration of water supply in India showed that personal bias — including "peer effects" — played a role.

On the plus side, the study found that stated satisfaction with the duration of water supply generally reflected the actual availability of water – i.e. satisfaction tended to increase with the hours per day that water was available. However, factors other than actual service provider performance did play a role. In particular, peer effects, or how service compares with that of their peers, had positive, and in some cases significant, effects. For example, results in Bangalore showed that going from about one-third of the number of hours of the reference group to an equal number of hours increased the probability of being satisfied with the service by 6% to 18% (depending on how you define the reference group). However, increasing actual water availability by one hour per day increased the probability of being satisfied by only about 1%. As the authors conclude, "an important policy implication is that overall satisfaction is to some extent a function of equality of service access. Everything else being equal, households will be more satisfied if service levels do not deviate significantly from those of their reference groups. Investment could thus be targeted specifically at reducing unequal service access by bringing the worst off neighborhoods up to the level of their peers."

50. Deichmann and Lall (2003).

Data from India shows an inverse relationship between self-rated health and health outcomes, showing that social context matters.

Similarly, one problem with relying on self-rated health is that people's perceptions are conditioned by their social context. For instance, a person brought up in a relatively "sick" community might think that his or her symptoms are "normal" when in fact they are clinically preventable. In contrast, a person from a relatively "healthy" community might regard a wider array of symptoms and conditions as indicative of poor health, regardless of how strongly these conditions are actually related to mortality. Economist and philosopher Amartya Sen argues that his may help explain why the Indian state of Kerala has the highest rates of longevity but also the highest rate of reported morbidity, while Bihar, a state with low longevity, has some of the lowest rates of reported morbidity. In such cases, providing good baseline information may help mitigate such peer effects.

Contribute

www.smartsummit_url.com

Importance of information

Improved access to information has resulted in a dramatic increase in the predictive strength of self-rated health between 1980–2002.

Access to timely, relevant information has also been found to play an important role in influencing the impact of feedback-related initiatives. For instance, yet another problem with relying on patient perceptions of health (in addition to peer effects as described above) is that while health is multidimensional, it is not entirely sensory. One study looking at the changing relationship between self-rated health and mortality in the U.S. between 1980 and 2002 found that the predictive validity of self-rated health increased dramatically during this period.[51] While the exact causal mechanism is unclear, the authors attribute this change to individuals' increased exposure to health information, not just from new sources like the internet but also from increasing contact with the health care system. Thus, access to information can put people in a better position to accurately evaluate the many relevant dimensions of health.

As with self-rated health, people's ability to adequately assess the quality of service provision depends, at least in part, on their access to relevant information. As mentioned above, the social accountability movement is in large part based on the hypothesis that a lack of information constrains a community's ability to hold providers to account. In one RCT, researchers tested two treatment arms: a "participation-only" arm and a "participation plus information" arm. The "participation-only" group involved a series of facilitated meetings between community members and health facility staff that encouraged them to develop a shared view of how to improve service delivery and monitor health provision. The "participation plus information" group mirrored the participation intervention with one exception: facilitators provided participants with a report card containing quantitative data on the performance of the health provider, both in absolute terms and relative to other providers. The "participation-only" group had little to no impact on health workers' behavior or the quality of health care while the "participation plus information" group achieved significant improvements in both the short and longer run, including a 33% drop in under-five child mortality.

According to one study, citizen participation in the monitoring of health providers had no impact on health outcomes when not accompanied by access to relevant information.

51. Schnittker and Bacak (2014).

When looking at why information played such a key role, investigators found that information influenced the types of actions that were discussed and agreed upon during the joint meetings – namely, the "participation only" group mostly identified issues largely outside the control of health workers (i.e. more funding) while the "participation plus information" group focused almost exclusively on local issues, which either the health workers or the users could address themselves (i.e. absenteeism).

Contribute

www.smartsummit_url.com

Level of technical difficulty

Other studies show that information has its limits too. In one of the few studies looking at the empirical relationship between corruption perceptions and reality, Benjamin Olken finds a weak relationship between Indonesian villagers' stated beliefs about the likelihood of corruption in a road-building project in their village with actual levels of corruption. Although villagers were sophisticated enough to pick up on missing expenditures – and were even able to distinguish between general levels of corruption in the village and corruption in the particular road project – the magnitude of the correlation was small.

A study in Indonesia found that bottom-up monitoring of corruption in a village-level infrastructure project had no impact on corruption while a top-down government audit reduced corruption by 8 percentage points.

Olken attributes this weak correlation in part to the fact that officials have multiple methods of hiding corruption and choose to hide corruption in the places it is hardest for villages to detect. In particular, Olken's analysis shows that villagers were able to detect marked-up prices but appeared unable to detect inflated quantities of materials used in the road project (something that arguably requires more technical expertise). Consistent with this, the vast majority of corruption in the project occurred by inflating quantities with almost no markup of prices on average. He argues that this is one of the reasons why increasing grassroots participation in the monitoring process yielded little overall impact whereas announcing an increased probability of a government audit (a more top-down measure) reduced missing expenditures by eight percentage points. These findings suggest that while it is possible that providing villagers with more information (in the form of comparison data with similar projects) could have improved the strength of the correlation between villagers' perceptions and actual corruption, information has its limits too: sometimes you just need experts to do the digging.

52. Olken (2006). Olken constructs his corruption measure by comparing the project's official expenditure reports with an independent estimate of the prices and quantities of inputs used in construction.

Contribute

www.smartsummit_url.com

Contribute

www.smartsummit_url.com

Section Four

CONCLUSION AND WAY FORWARD

In this section, we return to our original question: Is feedback the smart thing?

Our analysis suggests that – when properly implemented – feedback can be the smart thing. While only a handful of studies provide any direct evidence of impact, numerous other studies provide strong indirect evidence: namely, when asked to subjectively assess their own condition people consider a more inclusive set of factors than is otherwise possible to capture with more objective metrics. Second, people are not only good judges of the quality of services they receive, they can also pick up on important aspects of services that would otherwise go unmeasured but that may directly impact outcomes. Finally, even if our analysis were to show that constituent feedback was not a reliable source of information for policymakers, the feedback process itself can build trust and legitimacy, often a necessary condition for people to adopt interventions.

However, in stating our claim that feedback can be the smart thing, we are making a number of important assumptions. First, the entity (donor or government) on the receiving end of the feedback is both willing and able to act on it. Second, the people on the receiving end of services are active and willing participants in the feedback process and effective mechanisms exist to translate their preferences into actual outcomes. We know from numerous studies that this is not always – indeed rarely – the case.

Moreover, even if we make the assumption that feedback loops actually close, a number of additional demand-side factors – namely, personal bias, access to relevant information and technical know-how (or lack thereof) – may still get in the way of feedback being the smart thing. Here, the evidence suggests that the utility of feedback is largely incremental – not a perfect substitute for more objective measures – and likely to be enhanced when (1) complemented with access to relevant information, (2) well-designed and focused on user experiences (as opposed to technical aspects of service provision) and (3) adjusted for possible bias (through statistical tools or effective benchmarking to minimize peer effects), all of which take time, resources, and a deep knowledge of local context.

A number of key issues emerge as fruitful for future research.

- First, given that the evidence is still catching up to practice, we need more empirical studies using a variety of different research methods, from RCTs to case studies, to unpack when and how feedback improves outcomes. One challenge in building this evidence base is that, to the extent that successful feedback initiatives require a broader – or more "strategic" – package of reforms, an obvious question is how

we can adopt such an approach while at the same time isolate the impact of the feedback component itself. It seems that – if done right – most feedback initiatives would not lend themselves to RCTs, which try to avoid bundling interventions with other actions precisely for this reason. As an alternative, Jonathan Fox advocates for a more "causal chain" approach, which tries to unpack dynamics of change that involve multiple actors and stages.[54] A more nuanced discussion of what that would actually look like for a feedback initiative could be a useful endeavor.

- Second, in building the evidence base, we need to pay particular attention to how feedback compares to the next best alternative, including more top-down approaches. Most impact studies compare communities with the feedback intervention to control communities with no intervention (i.e. where the status quo is maintained). Few studies compare the feedback intervention to an alternate type of intervention that could help inform design.

- Third, we need to explore different incentives and mechanisms – both on the supply and demand sides – for "closing the loop." On the demand side, how do we ensure that participation is broad and that feedback is effectively aggregated and represented? On the supply side, what makes donors and/or governments listen – and how will we know if/when they actually do (as opposed to just "going through the motions" or checking off a box)?

- Fourth, to enhance the utility of feedback to policymakers, we need to test different ways of minimizing bias and better understanding the role that information plays in empowering people. Moreover, when does a lack of information cross over into being "too complex" for ordinary people to discern (and who gets to decide)? Put differently, which aspects of service provision are better left to people vs. experts to judge? And which combination of expert vs. constituent feedback produces the best results (i.e. the "sweet spot")?

- Last, as our analysis shows, feedback – if properly implemented – is not easy or free – it takes precious time and resources. We recognize that for cash-strapped donors and governments, answering the question "Is feedback the smart thing?" is not only about whether it leads to better social and economic outcomes but also whether it is cost-effective at scale. We found little guidance on this question in the empirical literature.

54. Fox (2014).

References

Andrews, Matt, Lant Pritchett and Michael Woolcock. (2012). *Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA)* (Working Paper 299). Washington, DC: Center for Global Development. http://www.cgdev.org/publication/escaping-capability-traps-through-problem-driven-iterative-adaptation-pdia-working-paper

Chris Argyris. (1991). *Teaching Smart People How to Learn.* Harvard Business Review. https://hbr.org/1991/05/teaching-smart-people-how-to-learn

Barder, Owen. (2009). *Beyond Planning: Markets and Networks for Better Aid* (Working Paper 185). Washington, DC: Center for Global Development. http://www.cgdev.org/publication/beyond-planning-markets-and-networks-better-aid-working-paper-185

Benyamini, Yael. (2011). *Why does self-rated health predict mortality? An update on current knowledge and a research agenda for psychologists.* Psychology & Health 26:11, 1407-1413. http://www.tandfonline.com/doi/pdf/10.1080/08870446.2011.621703

Björkman, Martina and Jakob Svensson. (2009). *Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda.* The Quarterly Journal of Economics. https://www.povertyactionlab.org/sites/default/files/publications/96%20Community-based%20monitoring%20of%20primary%20healthcare%20providers%20in%20Uganda%20Project.pdf

Björkman, Martina and Jakob Svensson. (2007). *Power to the People: Evidence From a Randomized Field Experiment on Community-Based Monitoring in Uganda.* The Quarterly Journal of Economics (2009) 124 (2): 735-769. https://www.povertyactionlab.org/sites/default/files/publications/96%20Community-based%20monitoring%20of%20primary%20healthcare%20providers%20in%20Uganda%20Project.pdf

Björkman, Martina, Damien Walque and Jakob Svensson. (2014). *Information Is Power: Experimental Evidence on the Long-Run Impact of Community Based Monitoring* (Policy Research Working Paper 7015). Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/2014/08/20144947/information-power-experimental-evidence-long-run-impact-community-based-monitoring

Boring, Anne, Kellie Ottoboni and Philip Stark. (2016). *Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness.* ScienceOpen Research. https://www.scienceopen.com/document/vid/818d8ec0-5908-47d8-86b4-5dc38f04b23e

Deichmann, Uwe and Somik Lall. (2003). *Are You Satisfied? Citizen Feedback and Delivery of Urban Services* (Policy Research Working Paper 3070). Washington, DC: World Bank Group. http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-3070

Donchev, Dilyan and Gergely Ujhelyi. (2013). *What Do Corruption Indices Measure?* Houston, Texas: University of Houston. http://www.uh.edu/~gujhelyi/corrmeasures.pdf

Gaventa, John and Rosemary McGee. (2013). *The Impact of Transparency and Accountability Initiatives.* Development Policy Review, 31: s3–s28. http://www.transparency-initiative.org/wp-content/uploads/2011/05/synthesis_report_final1.pdf

Glickman, Seth, William Boulding, Matthew Manary, Richard Staelin, Matthew Roe, Robert J. Wolosin, E. Magnus Ohman, Eric D. Peterson, and Kevin A. Schulman. (2010). *Patient Satisfaction and Its Relationship With Clinical Quality and Inpatient Mortality in Acute Myocardial Infarction.* Cardiovascular Quality and Outcomes (3: 188-195). http://circoutcomes.ahajournals.org/content/3/2/188.long

Hannan, Corinne, Michael J. Lambert, Cory Harmon, Stevan Lars Nielsen, David W. Smart, Kenichi Shimokawa and Scott W. Sutton. (2005). *A Lab Test and Algorithms for Identifying Clients at Risk for Treatment Failure.* Journal of Clinical Psychology 61(2):155-63. https://www.researchgate.net/publication/8121425_A_lab_test_and_algorithms_for_identifying_cases_at_risk_for_treatment_failure

Hayek, Friedrich A. (1945). *The Use of Knowledge in Society*. American Economic Review. http://www.econlib.org/library/Essays/hykKnw1.html

Fox, Jonathan. (2014). *Social Accountability: What Does the Evidence Really Say?* (GPSA Working Paper No. 1). Washington, DC: World Bank. http://gpsaknowledge.org/wp-content/uploads/2014/09/Social-Accountability-What-Does-Evidence-Really-Say-GPSA-Working-Paper-1.pdf

Kramer, Larry. (2004). *The People Themselves:* Popular Constitutionalism and Judicial Review. NY, NY: Oxford University Press.

Lambert, Michael J, Jason Whipple, David Smart, David Vermeersch, Stevan Lars Nielson and Eric Hawkins. (2001). *The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced?* Psychotherary Research 11(1):49-68. http://tartarus.ed.utah.edu/users/dan.woltz/EDPS%207400/Assignment%20Materials/Lambert_2001.pdf

Lambert, Michael. (2010). *Yes, It is time for clinicians to routinely monitor treatment outcome.* In B.L. Duncan, S.D. Miller, B.E. Wampold, & M.A. Hubble (Eds.), The heart and soul of change (2nd ed., pp. 239–268). Washington, DC: American Psychological Association. https://moodle.unitec.ac.nz/pluginfile.php/320563/mod_resource/content/1/Lambert,%20M.%20(2010)%20Yes,%20its%20time%20for%20clinicans%20to%20routinely%20monitor%20Treatment%20Outcome.pdf

Lindblom, Charles. (1959). *The Science of 'Muddling Through.'* Public Administration Review, 19, 79-88. http://www.jstor.org/stable/973677?seq=1#page_scan_tab_contents

Mansuri, Ghazala and Vijayendra Rao. (2013). *Localizing Development: Does Participation Work?* (Policy Research Report). Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/11859

Mason, Douglas, Mario Baudoin, Hans Kammerbauer and Zulema Lehm. (2010). *Co-Management of National Protected Areas: Lessons Learned from Bolivia*. Journal of Sustainable Forestry, 29: 2-4, 403-431.

Measures in Effective Teaching (MET) Project. (2013). *Ensuring Fair and Reliable Measures of Effective Teaching* (Policy and Practice Brief). http://files.eric.ed.gov/fulltext/ED540958.pdf

Measures in Effective Teaching (MET) Project. (2010). *Learning about Teaching: Initial Findings from Measures of Effective Teaching Project* (Research Paper). https://docs.gatesfoundation.org/Documents/preliminary-findings-research-paper.pdf

Minami, Tak and Jeb Brown. (N/A). *Feedback Informed Treatment: An Evidence-Based Practice.* (Blog post on A Collaborative Outcomes Resource Network (ACORN) website: https://psychoutcomes.org/COMMONS/OutcomesInformedCare.

Ostrom, Elinor. (2001). *Aid, Incentives and Sustainability: An Institutional Analysis of Development Cooperation* (Sida Studies in Evaluation 02/01). Stockholm, Sweden: Swedish International Development Cooperation Agency. http://www.oecd.org/derec/sweden/37356956.pdf

Olken, Benjamin. (2006). *Corruption Perceptions vs. Corruption Reality* (NBER Working Paper No. 12428). Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w12428

Olken, Benjamin. (2005). *Monitoring Corruption: Evidence from a Field Experiment in Indonesia* (NBER Working Paper No. 11753). Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w11753

Peixoto, Tiago and Jonathan Fox. (2016). *When Does ICT-Enabled Citizen Voice Lead to Government Responsiveness?* (Background Paper, 2016 World Development Report). Washington, DC: World Bank Group (in collaboration with Digital Engagement). https://openknowledge.worldbank.org/bitstream/handle/10986/23650/WDR16-BP-When-Does-ICT-Enabled-Citizen-Voice-Peixoto-Fox.pdf

Ramalingam, Ben. (2013). *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World.* NY, NY: Oxford University Press.

Schnittker, Jason and Valerio Bacak. (2014). *The Increasing Predictive Validity of Self-Rated Health.* PLoS ONE 9(1): e84933. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3899056/pdf/pone.0084933.pdf

Sen, Amartya. (2002). *Health: Perception versus observation*. BMJ 324(7342): 860–861. https://www.researchgate.net/publication/11415710_Health_Perception_Versus_Observation

Senge, Peter. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Doubleday/Currency.

Stern, Marc. (2010). *Payoffs versus Process: Expanding the Paradigm for Park/People Studies Beyond Economic Rationality.* Journal of Sustainable Forestry, 29: 174-201. https://www.researchgate.net/profile/Marc_Stern2/publication/233144192_Payoffs_Versus_Process_Expanding_the_Paradigm_for_ParkPeople_Studies_Beyond_Economic_Rationality/links/552bc7890cf21acb091e6f2f.pdf

Stewart, Moira, Judith Belle Brown, Allan Donner, Julian Oates, Wayne W. Weston and John Jordan. (2000). The Impact of Patient-Centered Care on Outcomes. *Journal of Family Practice* 49(09):796-804. http://www.jfponline.com/home/article/the-impact-of-patient-centered-care-on-outcomes/78c6a0031eb6ef3aae1e31851a4b8327.html

Tyler, Tom. (1990). *Why People Obey the Law.* New Haven and London: Yale University Press.

World Bank. (2004). *World Development Report 2004: Making Services Work For Poor People.* Washington, DC: World Bank Group. https://openknowledge.worldbank.org/bitstream/handle/10986/5986/WDR%202004%20-%20English.pdf?sequence=1

Zeitlin, Andrew, Abigail Barr, Lawrence Bategeka, Madina Guloba, Ibrahim Kasirye, Frederick Mugisha and Pieter Serneels. (2010). *Management and Motivation in Ugandan Primary Schools: Impact Evaluation Final Report.* Oxford, UK: Centre for the Study of African Economies, University of Oxford. http://www.iig.ox.ac.uk/output/reports/pdfs/iiG-D10-UgandaPrimarySchoolsImpactReportFinal.pdf