What have we learned here? Questioning accountability in aid policy and practice

Lauren Kogen

Abstract: In recent decades, development donors in the West have touted a shift to rigorous evaluations and evidence-based policymaking in order to address global skepticism regarding the effectiveness of aid. In the accompanying rhetoric, "accountability" and "learning" have been held up as twin pillars that will ensure a more effective aid-making system. This paper questions the ability of these concepts to improve aid in their current working forms. The paper offers a revised conceptualization of learning in order to improve funding and funding policy. The revised definition supports two particular areas in which "learning" is sorely needed but which are eschewed in most current institutionalized evaluation rhetoric: developing theory undergirding social change (such as theory behind gender-based violence) and evaluating project design and implementation processes (such as participatory designs).

Keywords: Communication for development, communication for social change, monitoring and evaluation, foreign aid

Introduction

Despite decades of practice, we in the West do not fully understand how to 'do' international development. Many development efforts fail, and it is difficult to predict which will succeed (Glennie and Sumner, 2014). Some would even say that after more than half a century's worth of efforts, we have done more harm than good (Easterly, 2006; Moyo, 2009). This is why we are still asking such fundamental questions as 'does aid work?' (Burnside and Dollar, 2004).

In recognition of the fact that we do not have all the answers, that many aid projects have failed, and that there is a global skepticism about the effectiveness of aid, policymakers have touted a shift to "evidence-based policymaking" in which research and evaluations serve to improve aid-making (e.g., USAID, 2016). Project evaluations have thus received much greater attention in recent decades, and aim to improve development efforts by prioritizing 1) accountability and 2) learning. Indeed, these two goals are cited by USAID (2016: 8) as the "two primary purposes" of evaluation.

Accountability typically refers to the question of whether or not a project met the goals it set out to achieve. In order for funders and implementers to be held accountable to taxpayers and beneficiaries, the logic goes, they must be able to show that they are spending money on development programs that are working and not wasting money on those that are not. *Learning* is, in theory, achieved by using data from past projects to improve future projects.

However, this paper argues that these terms have been used too close to interchangeably by the world's major donors. While accountability is relatively well established as a concept, learning, as a cross-cutting conceptual goal of evaluations, is relatively watery. This has severe

implications for what we can hope to achieve in terms of using evaluations to improve future interventions.

What is more, while Western donors still lack a fundamental understanding of what works in aid, 'learning' as currently conceptualized does not aim to answer such fundamental questions. Despite an acknowledgement that we still have much to learn when it comes to aid, present practices reinforce the assumption that our strategies are working reasonably well and that evaluations can, perhaps, provide some fine-tuning (Power, 1997). This means that using current definitions of accountability and learning as starting points for evaluations is pushing us away from answering foundational questions about how aid works, and ultimately making it harder, not easier, to improve development projects.

Much recent work has focused on the idea that accountability- and learning-based approaches are in many ways incompatible (Armytage, 2011). I expand here upon this idea and argue, first, that accountability and learning have been inappropriately conflated in the documentation of two of the world's major donors – USAID and DFID – stemming from the assumption that accountability, by definition, improves projects. This conflation, in turn, makes it much more difficult for project implementers and evaluators to understand which learning-based questions policymakers need answered in order to improve aid priorities. Second, I argue that a revised conceptualization of "learning," focused on theories undergirding social change and social change project implementation, would make evaluations more useful..

My focus here is on Communication for Development and Social Change (CDSC) projects. CDSC projects, as defined by Lennie and Tacchi (2014), focus on social change and tend to deal with interpersonal or other forms of communication between individuals, groups, and / or institutions. In this sense the field is very broad; theory related to CDSC design and

evaluation speaks to all development projects that focus on human-centered social change – the kind of complex, messy scenarios that make interventions and evaluations so complicated, that lend themselves to failure, and for which making assumptions about what works is particularly dangerous.

Current conceptualizations of accountability and learning by USAID, DFID, and OECD

Placing accountability at the forefront of evaluations is the guiding principle behind results-based management (RBM), the preferred management style of many of the world's major donors (Vähämäki et al, 2011). RBM rose in prominence following the Paris and Accra High Level Forums on Aid Effectiveness convened by the Organization for Economic Cooperation and Development (OECD) in 2005 and 2008, to which 138 countries agreed to adhere. Following increasing global skepticism about the effectiveness of aid, this series of forums take as their raison d'être the notion that "aid effectiveness must increase significantly" (OECD, 2008) and that effectiveness must be measurable and visible to the public. The Paris and Accra meetings resulted in a set of principles that together are assumed to make aid more effective. One of these key principles is "managing for results": the need to "manag[e] and implement[] aid in a way that focuses on the desired results…" (OECD, 2008: 7). The Evaluation Policy Document produced by the U.S. Agency for International Development (USAID) follows suit and cites the OECD's principle of managing for results, stating that projects must monitor "whether expected results are occurring" (2016: 9). The organizations rightly acknowledge that it is unethical and nonsensical to oversee billions of dollars of aid without knowing whether or not that aid is doing any good. They conclude that measuring levels of success will ensure that only projects that are truly improving lives will be funded and supported. In this way, the organizations become

"accountable" to those the projects are aiming to help, because RBM ensures that money is not being wasted, or failing to reaching beneficiaries, or simply going to Western-based staff for projects that are failing. RBM also promotes accountability to taxpayers, who are ultimately supporting the bulk of these projects.

Likewise, these large donor organizations tout the importance of *learning* from past interventions in order to improve future interventions. Donors expect evaluation questions to be "explicitly link[ed]" to policymaking decisions (USAID, 2011: 7) and support evaluations that produce "high quality evidence for learning" (DFID, 2014: 3).

These conceptualizations of accountability and learning seem logical and straightforward at first glance, but they require parsing. There are thousands of development projects currently underway. What is the best way to take all of these projects and "learn" from them? What precisely are we trying to learn? How do we extract lessons that will make development aid work better for beneficiaries? The evaluation documentation published by USAID and the U.K.'s Department for International Development (DFID) (the world's largest foreign aid donors) do not explicitly address this. Rather, a close reading of the documentation reveals two problematic assumptions about how learning results from an RBM approach: first, that accountability and learning are essentially equivalent, and second, that periodic literature reviews help extract lessons across a portfolio of evaluations. I question these assumptions below and suggest that this view of learning is short-sighted and does not produce high-quality learning.

Conflation of accountability and learning

The policy documents produced by USAID and DFID eschew any clear definition of learning. For example, USAID's Evaluation Policy Document defines learning as

"systematically generat[ing] knowledge about the magnitude and determinants of project performance" (2011: 3). On its face, it is difficult to differentiate this definition from a common understanding of accountability, save for the word 'determinants,' which implies some attention to why a project succeeded or failed. DFID's 2014 Evaluation Strategy gives no guidance at all on how it views learning even though, as in the USAID documentation, it emphasizes the importance of learning. Both documents incorporate a general sentiment that learning should improve future program planning, but provide minimal information regarding what the organizations hope to learn or what type of data would be most useful for learning. The OECD's Principles document hardly mentions learning, except to say that lessons learned should be shared (2008: 6). This lack of a clear definition leaves the term open to interpretation, with worrisome implications for improving the effectiveness of aid.

One implication of the repetitive reference to the importance of "accountability and learning" without a formal definition of learning is the suggestion that accountability and learning are two sides of the same coin: that by addressing one we are necessarily addressing the other. Indeed, when learning *is* referenced in evaluation documentation, the general impression that arises is that learning simply refers to knowing whether projects worked or not and to what degree. But this is precisely the aim of accountability. Accountability is about understanding whether a project has succeeded or failed; learning here suggests that by seeing which projects have succeeded and failed one can *learn* which projects to promote and which to terminate. The implication thus becomes that learning is solely about assessing project success and that it is therefore a natural outcome of accountability. This is the implication of statements such as USAID's claim that learning "represents a continuous effort to… measure the impact on the

objectives (results) defined" (2011: 3) or DFID's statement that "monitoring results provides us with an incentive to look at the evidence, innovate and learn" (2014: 1).

Neither explain what we might learn beyond whether or not to re-fund or de-fund the project at hand. But "learning" whether or not an individual project worked does not necessarily help us understand development as a field. The most that can be done with this form of learning is to a) keep funding the same project, or b) fund an identical project in another location. The dangers of this understanding of learning should be obvious: We cannot improve interventions by copying and pasting. Understanding that intervention A works to improve problem X is much less useful than understanding, for example, what kinds of broad-based strategies work to improve problem X, and why, under what circumstances, etc. *This* understanding of learning would do much more to help practitioners and policymakers improve the world of development interventions.

A popular recipe website, allrecipes.com, features 885 different recipes for chicken parmesan. One might determine that all recipes with four or more stars (out of five) should be considered successful and that all other chicken parmesan recipes should be discarded. However, this criterion still leaves us with hundreds of "successful" recipes. Furthermore, someone who is trying to extract lessons, to teach other chefs how to go about making a successful chicken parmesan, should be able to do more than advise chefs to make one of the 300 or so recipes that happen to appear on the website. They should be able to provide overall lessons about the best ingredients and the best cooking techniques. Likewise, policymakers should not be taking an allrecipes.com approach to aid-making: They should be learning what works overall – general principles about how to conduct effective aid-making. If asked what makes for an effective HIV prevention strategy, for example, a well-informed policymaker should be able to do more than

advise practitioners to copy one of the hundreds of HIV interventions that have already been conducted.

There are no fool-proof recipes for development. Someone who wants to prevent HIV in New Delhi cannot pluck an intervention plan that was used in rural South Africa from a database, insert it into a vastly different context, and expect it to work in the same way. General guidelines, principles, and theories are needed so that interventions can be designed for the various contexts in which they are implemented. In other words, learning *why* particular programs work in particular contexts is what contributes to improving policy, not simply learning *whether* programs worked, and this question is all too frequently left out of accountability-based evaluations.

Indeed, Deaton (2009: 4) advocates the formation of "potentially generalizable mechanisms" that help explain why a project worked. Likewise, Stern et al. (2012: 27) argue that project findings need to "form[] the basis for more general theory" in order to streamline future projects. This is particularly crucial for projects that address human behavior and social change in which we struggle to understand what leads individuals and institutions to change their thoughts and behaviors and in which well-designed interventions often do not go according to plan. Servaes (2016: 2) notes that 75% of recent CDSC work has been atheoretical – referencing no theory at all – and that of those pieces that do cite theory, outdates theories stemming from the modernization paradigm are the most common.

A project that provides food to schoolchildren, for example, might succeed in increasing school attendance, but without learning how parents make decisions about when to send their children to school and when to keep them at home, funders cannot extract lessons about what would have made the project even more successful, whether the project could be replicated in

another setting, or whether the project is sustainable. It may be the case that in certain contexts the cost of sending children to school (the danger and length of the commute, the loss in wages, etc.) outweighs the benefits of an additional meal. High quality evaluations may incorporate this kind of learning, but a lack of emphasis on this kind of learning means that many evaluations do not. It is in part for this reason that many evaluation experts describe "accountability-based" and "learning-based" evaluations as at odds with each other (Armytage, 2011; Cracknell, 2000; Lennie and Tacchi, 2014), despite the insistence by organizations like USAID that both concepts are fundamental and can be "mutually reinforcing" (2016: 8). Lennie and Tacchi (2015: 27) define learning-based approaches as those that "understand social change as emergent, unpredictable, unknowable in advance." In other words, a focus on learning assumes that we are learning about *how social change works*, not about whether *a particular project worked*.

There is another concern with conflating accountability and learning and using accountability (whether or not a project succeeded) to determine policy (which projects should be supported). Let us assume, for a moment, that donors should continue funding all successful projects. Not only does this not tell funders what makes particular projects successful, but it means that they are not being very selective in the projects they are funding. This is because almost all human-centered interventions 'succeed'. If we define 'positive impact' or 'success' as achieving the goals the project set out to accomplish, it does not take much for a program to succeed, as long as the organization is savvy and does not overestimate its potential impact. RBM emphasizes quantitative "indicators" and "targets" and "log frames" to ensure projects are truly achieving measurable objectives – but the very results-based system that was implemented to ensure accountability has led to a system in which program designers are forced to reduce goals to simple, straightforward, low-level, moderately-easy-to-achieve results (Armytage, 2011:

68; Gumucio-Dagron, 2009). They do this, first, so that they are able to translate complex project goals into a number-based spreadsheet that requires everything to be broken down into their component parts, and secondly so that organizations are seen as performing well when it comes time for final evaluations. Such evaluation mindsets work better for public relations priorities than for furthering global development (Enghel, 2016).

Major funders like USAID and DFID should be able to do better than simply improving semi-successful projects. They should be using evaluations to actively decide which types of activities should be prioritized and which have the highest potential to significantly improve lives. But USAID explicitly acknowledges that simple improvement to existing projects is their primary objective. Its documentation suggests that *the* purpose of learning is to allow "those who design and implement projects, and who develop programs and strategies… to refine designs and introduce improvements into future efforts" (2011: 3), but learning should mean much more than refining a single project. Evaluations that explain why certain projects succeed or fail and provide recommendations for overall policies can help refine funders' understanding of development and social change and help improve policy-making decisions. Understanding *why* a project works or fails and tying these findings to policy recommendations therefore needs to be intimately tied to the priorities of evaluations and to the working definition of learning.

Literature reviews as panacea

The evaluation documentation of USAID and DFID does reference the idea of understanding why a project succeeded or failed (e.g., USAID, 2016: 8), but rarely. Instead, the documentation suggests that real learning, the ability to understand what works across contexts, comes from reviews: periodic syntheses of research, reports, evaluations, etc. that allow

policymakers, scholars, and practitioners to produce overall lessons accumulated after analyzing many projects. Through such a process, it is suggested, patterns will emerge, and *learning* will ensue.

Many of these agencies commission literature reviews for this very end. USAID commissions "systematic reviews" to "synthesize the available research on a specific question" (2016: 20); DFID similarly assumes that its learning will take place through "thematic [and] sectoral synthesis" of findings (2014: 13). Extending from the assumption above that a proper RBM approach will necessarily lead to learning, the idea here is that looking across a portfolio of projects that have succeeded and failed will provide additional insights into what kinds of approaches succeed or fail.

But this is a very limited version of learning for two reasons. First, if evaluations focus on accountability, or success, of a project, rather than an understanding of why the project succeeded or failed, there is no reason to believe that looking at many evaluations will bring to light new lessons that did not appear in the evaluations of individual projects. Of course, it is possible that patterns will jump out. It is possible that the vast majority of four-star chicken parmesan recipes, for example, use parmesan cheese and San Marzano tomatoes. If this were the case, a perceptive reviewer could conclude that these ingredients should be considered default strategies for making chicken parmesan.

Alas, lessons about development are rarely so obvious. The "ingredients" that go into projects are vastly more complex, and those ingredients interact with the setting in which projects are implemented. Even if there is meticulous monitoring of project inputs, this is usually not enough to help explain why a complicated human-centered project, such as encouraging

people to use condoms, succeeds or fails. Without the why, the lessons extracted from a review are very limited.

This is exacerbated by the fact that reviewers find it very difficult to deal with disparate data collection methods when trying to synthesize large numbers of studies. Qualitative data is particularly problematic, yet qualitative data is often how we best answer questions of "why." Thus, many evaluations that rely on qualitative data are excluded from literature reviews. For one reason or another, they are designated as unhelpful by reviewers. There are a variety of reasons for which reviewers may exclude studies – that the methods are flawed, the location is mismatched to the needs of the reviewer, the questions asked are not pertinent, etc. – but ultimately many studies are eliminated because they are not conducted in a way that helps reviewers synthesize or extract lessons. USAID explicitly states that its systematic reviews are typically only used for randomized controlled trials (USAID, n.d.), which by their very design emphasize answering whether a project worked rather than why (Deaton, 2009) and tend to emphasize quantitative data.

This is not to say that literature reviews are not useful, or that patterns do not emerge. Sometimes literature reviews reveal important patterns. But reviewers are neither clairvoyants nor magicians; they cannot see, nor draw conclusions about, what is not suggested or addressed in individual reports. If evaluations are focused exclusively on whether or not projects work, it is too easy for evaluations to simply become a series of case studies – disconnected project reports that hint at larger lessons about aid but make it difficult for reviewers to connect the dots. It is often argued that evaluators can make connecting the dots easier for those conducting literature reviews by using common indicators (numbers), but another, I would argue more promising

strategy, would be to make qualitative findings more meaningful and useful. I discuss ways to accomplish this in the next section.

It is not only literature reviewers that have trouble using evaluations and other studies of development efforts to improve general understanding of development and development policy. A study conducted by Kogen et al (2012) that surveyed and interviewed policymakers and practitioners working in communication and development found that overall, policymakers felt that the available literature was impractical. They stated a preference for "research findings [that] are translated into clearly–articulated [*sic*] and replicable recommendations" (2012: 3). In survey responses, policymakers scored the "usefulness" of existing evidence relating to the role of media and communication in development at a mere 3.2 out of 5 (close to neutral) and, on average, said they "sometimes" consult it. This does not bode well for a field such as CDSC that is trying to achieve greater inroads into policymaking and convince policymakers of its usefulness. (Indeed, many policymakers are not even sure what CDSC *is* (Feek and Morry, 2009; Lennie and Tacchi, 2015).) The inattention to larger policy recommendations is not limited to CDSC work. Even when it comes to randomized controlled trials in various fields of development, policymakers often find results impractical, and want answers beyond whether or not an intervention worked (Ravallion, 2009: 2).

Re-defining 'learning'

How, then, can we make learning from evaluations more meaningful and useful? The first step is to recognize that accountability and learning are not the same thing and should not be conflated. A loose use of the word learning is undergirded by a misplaced dependency on accountability and RBM as cure-alls for what ails development funding. Accountability, on its

own, does not improve aid, and yet it is clear that in rhetoric as well as practice that accountability is being prioritized above learning and that learning is seen as icing on the cake – a perk resulting from good accounting.

But learning is what improves lives, and for this reason learning must be prioritized (as others have also argued (e.g., Lennie and Tacchi, 2014)). Accountability does not easily contribute to a larger conversation about aid, and it does not directly benefit stakeholders, it only provides legitimacy for funders and practitioners to keep pursuing current development projects (Power, 1997). Furthermore, learning cannot simply be seen as an automatic byproduct of accountability, as learning requires "different data, methodologies, and incentives" (Armytage, 2011: 263).

To be sure, funders should be held accountable to their various stakeholders; they should not be funding projects that are failures. But measuring success should be viewed as a secondary purpose of evaluation – as a means to an end, in which the "end" is learning how to effectively promote development.

Meaningful lessons need to come about at the level of individual projects, not at the stage of cumulative literature reviews. In order to answer meaningful questions about what works in aid, looking at project success or failure is not the most productive starting point. Understanding whether a project worked is much simpler than trying to understand why it worked, and is much less useful. I suggest that a focus on "why" should be broken down into two broad fields in which there is currently a large gap in knowledge. The first regards theories about what promotes social change; the second regards understanding how to actually design and implement effective development interventions.

Proposed learning priority 1: Developing theory-based guidelines for social change

Even if donors and practitioners agree that answering the question of why and how social change occurs is important, it is still not straightforward how we ought to answer these questions. Imagine a fourth-grader who is working on his long division. If he gets an answer wrong, his teacher can encourage him to figure out exactly where he went wrong when solving the problem. But what if the student gets the question right? How could the student begin to answer the question of where he went… 'right'?

This question is not just rhetorical. For a project addressing HIV prevention, for example, an organization's evaluation of why a project was successful could be that the organization focused on addressing social stigma (because it determined this was a key barrier in preventing HIV) or it could be that it hired a top-notch staff that kept the project running efficiently and smoothly. One of these findings is very useful in terms of thinking about what works in HIV prevention; the other far less so.

And yet, project evaluations very often focus on these latter kinds of site-specific "why" lessons on how to improve a particular project, such as shifting staff responsibilities, re-allocating funds, focusing more on a particular activity, or adding a new activity. Learning "why" a project had particular outcomes or "how" it could be improved then emphasizes recommendations for project refinement, rather than providing lessons that can actually improve aid policies overall. USAID acknowledges this, stating that the most common ways that evaluations are used are to "refocus ongoing activities, including revisions to delivery mechanism work plans, extending activity timelines or expanding activity geographic areas" (2016: 16). Given how much we still need to learn about development, limiting learning efforts to how to improve individual projects is rather narrow.

Instead, learning should be focused on generalized theory about why and how social change occurs, and development of such theory should be considered a key piece of what is needed in order to improve aid effectiveness. This recommendation aligns with the calls of those who have advocated for increased attention to why projects work, such as Deaton (2009) and Stern et al (2012).

Evaluators therefore need to take this extra step of providing generalizable takeaways, not leave the broader lessons implicitly embedded in the project details for a perceptive literature reviewer to uncover.

Donors could move evaluators in this direction by prioritizing questions about particular genres of social change. These genres need to be broad enough that they do not presume causal pathways to development, but not so broad that they become meaningless. Questions like "How do we use mobile phones to improve government accountability?" or "What are the best methods for opening up media systems?" or "Which works better for persuasion – information or entertainment?" all make assumptions about what works in development. These sorts of questions have their place – and sometimes funders have reasons for asking specific questions. But these very precise questions should not take up the bulk of development funding because they have already missed the boat. They eliminate other potential pathways for change. They assume that we know how to 'do' development and that we are at the stage of refining our technique. We are not.

For example, there is currently much research on, and evaluation of, projects that seek to strengthen citizen voice and government accountability through information communication technologies (ICTs) like mobile phones. This work follows an effort by aid policymakers to pursue ICT-related development projects (such as the 'Making All Voices Count' development

program currently supported by various international governments). Yet, there is a dearth of evidence that increasing access to ICTs is an effective fix for improving governance and government accountability (Gagliardone et al., 2015; Gillwald, 2010). Indeed, Shah (2010: 17) argues that the obsession with ICTs echoes the modernization paradigm, in which new technologies are "accompanied by determined hope that Lerner's modernization model will increase growth and productivity and produce cosmopolitan citizens." Excitement about the latest innovations in development may be causing policymakers to jump the gun when it comes to choosing the most practical types of programs rather than asking basic questions about how to improve governance.

Questions emphasized by donors should directly address how to improve lives. If large stakeholders could come to an agreement on a handful of these questions, and funders, evaluators, and designers created projects and evaluations that addressed them, this would work to make these disparate, case study projects more comparable, and the job of synthesizing extant data much more productive.

One example of such a question, or genre, relating to social change could be how to reduce gender-based violence (GBV). If projects addressing GBV provided big picture, theory-based, actionable answers to this question, this could help push forward more effective GBV donor policy. For example, one theory about gender-based violence is that for many women, failure to leave abusive relationships is based in large part on financial insecurity, and that if women had their own source of income, this would help them become less financially dependent on men, and therefore less likely to be victims of abuse. If evaluations could provide evidence for policymakers about whether or not this theory is sound, why, and under what circumstances,

these insights could easily be incorporated into granting practice and future project designs across regions.

One might argue that this is already being done – that all evaluations of projects regarding GBV address (explicitly or not) broader lessons about how to reduce GBV. But this is precisely what I am suggesting is *not* the case as often as one might think, and that the degree to which these lessons are included is inadequate. As described above, the plethora of development evaluations have not led to as large a knowledge base as one would expect in particular development sub-fields. In the case of GBV, for example, the knowledge base "about effective initiatives is [still] relatively limited" (Bott et al., 2005: 3).

This is because many evaluations do not provide actionable insights on what improves lives. Some evaluations focus on whether or not an individual project worked (for instance reporting whether trauma from GBV was reduced by an intervention, without addressing why); some make assumptions about how to reduce GBV without providing sufficient basis for the assumption (for example assuming that GBV should be reduced through counseling by local health care providers); some only provide recommendations about how to make the specific project better rather than broader takeaway lessons about how to reduce GBV (for example by adding more local staff). While there are many evaluations and reports that do provide important insights, my sense from my own experience with the literature, project funders, and project practitioners, as well as the continuing lament of scholars that high quality, rigorous data regarding important development questions is lacking, suggest that this is not nearly done often enough.

Proposed learning priority 2: Social change process evaluation

Another major assumption of current evaluation policy is that, if we are implementing a project that is based on solid evidence about what promotes positive change, the process of actually designing and implementing the intervention in a particular context is straightforward. The fact that funders do not provide guidelines for or a specific mandate to evaluate the design and implementation processes suggests that this is the case. However, designing and implementing a development intervention is quite difficult, as anyone who has ever tried to implement a project in unfamiliar territory will attest. Teams need to work with locals, achieve community buy-in, and understand local hurdles to change, among other challenges. Certain project teams have ways of creatively and dynamically engaging with communities; other teams seem to get stuck early on, dealing with communication problems and complex foreign environments they cannot comprehend or negotiate. There are too many projects in which teams paid insufficient attention to local context, did not factor in local institutions or traditions, did not understand ways that the problem was already being addressed, or did not understand what the community truly needed in order to instigate change.

Understanding how to implement projects effectively is crucial, and yet for some reason this is not an area that is frequently evaluated or for which project reports and evaluations provide lessons that can inform the development field as a whole. Just because one project distributed menstrual pads to school girls and succeeded in increasing their school attendance does not mean that another organization that attempts the same thing, even in the same community, will be successful.

One major conversation happening right now around project design addresses participatory designs, in which beneficiaries are actively involved in the design, implementation, and even evaluation of projects (see, e.g., Lennie and Tacchi, 2013). The idea is certainly not

new, but recent conversation stems, in part, from an increased recognition that top-down designs that do not take into account feedback and advice from target beneficiaries are less ethical and less likely to succeed. That said, while many projects engage with their beneficiaries to some degree, at least for feedback on design (Huesca, 2003), it is rare for even so-called participatory projects to provide details about how participation was designed, encouraged, and achieved, or how 'participatory' a project actually was (Inagaki, 2007). So participatory design is at a stage in which many accept its legitimacy, but it is insufficiently clear how to successfully implement such a design, how many practitioners are using it and to what to degree, or how such interventions should be evaluated (Waisbord, 2015).

Therefore, with regard to design and implementation evaluation, one specific question evaluations might start answering would be whether deeply integrating stakeholders into the design and implementation process actually improves project outcomes. This is an assumption of much development work (and particularly CDSC work and projects that incorporate participatory elements) and the logic seems sound, but we do not have sufficient evidence that this is really the case. To be sure, the fact that so few practice it suggests that most believe it is *not*. There is an underlying normative implication in discussions of participatory research and evaluation that participatory designs are more ethical (Huesca, 2003), but we do not have sufficient evidence that participatory designs actually produce better development outcomes. In fact, of the limited literature on participatory designs, a significant portion addresses challenges in (e.g., Banerjee et al., 2010; Botes and van Rensburg, 2000) and some that have tried to implement them have run into problems on the ground (e.g., Campbell & MacPhail, 2002; Lennie & Tacchi, 2014: 16).

In theory, participatory design holds significant advantages over traditional outsider-led project design, in that it is more likely to match the needs of the community and less likely to suffer from major oversights since the community itself is involved. These assumptions have indeed been borne out in many participatory-based projects (e.g., Cashman et al., 2008). These factors would point to participatory projects having greater impact (because they more directly address the needs of the community and avoid missteps) and being more sustainable (because the community has more ownership and buy-in of the project). On the other hand, participatory design is time-consuming, potentially limited to small interventions with a small number of beneficiaries, and therefore potentially too inefficient to promote significant social change (Hornik, 1988). Having convincing evidence that participatory project design actually accomplishes development goals in a way that is practical and replicable could create a sea change in the world of development funding. This is an important area for which we need to be providing policymakers with some fundamental evidence that doing projects in a different way is worthwhile. Frameworks for tracking such processes have already been laid out by the evaluation community (Figueroa et al., 2002), but it is as yet unclear to what degree these schema for social change process evaluation have been utilized in the field.

Together, the two broad evaluation priorities defined above – developing theory-based guidelines for social change and evaluating design and implementation processes – can help begin to answer some basic questions about what is going 'right' in interventions that succeed and what is going wrong in those that do not. Both can be crucial for improving policy. Not every evaluation will address these two topics, not every project will align with this model, but more need to if we are to begin improving aid policy in a meaningful way.

Conclusion

I have argued that the twin pillars of today's development evaluations – accountability and learning – should not be conflated. The terms, as they are constructed by the world's largest donors, assume that donor communities possess a basic competence in development that is simply in need of proof to confirm its benefits, and perhaps some fine tuning. Rather, evaluations need to be reimagined in a way that help us answer more foundational questions about how to 'do' development.

As I have argued above, simply knowing whether or not a project was successful does not, on its own, do anybody any good, especially if most human-centered social change projects technically 'work.' When we pledge accountability to taxpayers and beneficiaries, this should mean more than providing accounting mechanisms. Being able to record the "amount" of impact we had on developing regions does not benefit anyone unless it contributes to increasing positive impacts down the line. The only way to increase positive impact is to properly learn from projects. The only way to properly learn from projects is to acknowledge that we do not know much, and then to prioritize learning above accounting. The attention to results and impact and the relegation to the back burner of methods for learning and improvement suggests that funders either do not truly believe they need to learn or are at such a loss that they do not even know how to articulate what it is they wish to learn. This is why answering whether a project "worked" is the wrong place to begin an evaluation.

William Easterly (2009) spends the bulk of his book *White Man's Burden* providing convincing arguments that the West does not know what it is doing when it comes to development, that it has had far more failures than successes, and that despite this it evokes a "patronizing confidence" that it knows how to solve development problems (2009: 368). A shift

in evaluation goals from accountability to learning would signal an acknowledgement that we have much to learn about the basics of aid. This shift would define "learning" as answering foundational questions about aid rather than fine tuning existing programs that themselves may be misguided.

Specifically, I suggest that "learning" focus on two broad areas: underlying theories about what promotes social change and the process of designing and implementing interventions. Theories that have not been confirmed on the ground should not serve as a backbone for large and ambitious projects, such as using ICTs to promote the economy or government accountability. By skipping the foundational evidence, we risk wasting precious aid dollars on projects that are based on assumptions rather than evidence.

Secondly, evidence on how to effectively design and implement interventions is lacking. Evaluations and reports infrequently analyze the process of conducting development interventions, suggesting that this is an area that does not require analysis because best practices are sufficiently established. They are not.

Many organizations do indeed incorporate useful learning into their evaluations. Good project reports and evaluations include recommendations that are broad and useful to those trying to understand, generally, what works in development. These do not always come in the form of evaluation reports. For example, BBC Media Action often issues "policy briefings" that distill findings from their interventions in order to make them easy to understand and be used by policymakers.

I have proposed specific questions that could be asked in each of these broad categories. With regard to underlying theories, I proposed deep and critical examination of the question of how to reduce gender-based violence. With regard to design and implementation processes, I

proposed explicit evaluation of participatory design, an approach to project development and implementation that has received increasing attention in the literature, seems to offer potential as being more effective and more ethical than traditional, top-down implementation designs, and yet it is mostly ignored (except to a superficial degree) by most funders and designers who want more control over the project. This may be, in part, because there is little evidence as of yet that participatory design actually produces better outcomes. The number of projects that fail to take realities on the ground into sufficient account, thus weakening the effectiveness of the project, suggests that this is a question that merits more attention.

Shifting the focus of evaluations away from impact may be anathema to project staff that want to tout the success of their projects. This is understandable. Organizations depend on positive evaluations to receive continued funding. Without evidence that they are doing good work, they risk losing the ability to continue working. But I have not argued here that evaluation reports should not address impact. What I have argued is that we need to change the mindset around evaluations – the ideology that evaluations are meant to assess impact, that this is their primary function, and that this is what makes development aid accountable and ethical. I have tried to argue that this frame is, in fact, *not* the most ethical and just, because it limits the usefulness of evaluations and of projects. It shows that money was not entirely wasted but it does not efficiently or sufficiently teach us better ways to spend money. In this way it benefits donors and implementers far more than the beneficiaries they purport to help. Shifting the focus of evaluations to understanding fundamental questions about how aid works may be a better, more ethical, and more just way to improve policy making and to stay accountable to beneficiaries at the same time.

We could simply be happy with the fact that our interventions are working, and continue to fund those that work and re-allocate the funds of those that do not. But we should not be satisfied with projects that simply work. We should be seeking to make the most efficient use of money possible and improving lives to the largest degree that we can. This is what we should mean when we refer to "accountability."

References

Armytage L (2011) Evaluating aid: An adolescent domain of practice. *Evaluation* 17(3): 261–276.

Banerjee AV, Banerji R, Duflo E, Glennerster R and Khemani S (2010) Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy* 2(1): 1–30.

Botes L and van Rensburg D (2000) Community participation in development: Nine plagues and twelve commandments. *Community Development Journal* 35(1): 41–58.

Bott S, Morrison A and Ellsburg M (2005) *Preventing and responding to gender-based violence in middle and low-income countries: A global review and analysis*. Policy research working paper no. 3618. Washington DC: World Bank.

Burnside C and Dollar D (2004) *Aid, policies, and growth: Revisiting the evidence*. Policy research paper no. O-2834. Washington DC: World Bank.

Campbell C and MacPhail C (2002) Peer education, gender and the development of critical consciousness: Participatory HIV prevention by South African youth. *Social Science and Medicine* 55(2): 331–45.

Cashman SB, Adeky S, Allen AJ III, Corburn J, Israel BA, Montaño J, Rafelito A, Rhodes SD, Swanston S, Wallerstein N and Eng E (2008). The power and the promise: Working with communities to analyze data, interpret findings, and get to outcomes. *American Journal of Public Health* 98(8): 1407-1417.

Cracknell B (2000) *Evaluating development aid: Issues, problems and solutions*. London: Sage.

Deaton A (2009) *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development*. Working paper 14690. Cambridge, MA: National Bureau of Economic Research.

DFID (2014) *DFID evaluation strategy 2014-2019*. London: Department for International Development.

Easterly W (2009) *The white man's burden: Why the West's efforts to aid the rest have done so much ill and so little good*. Oxford: Oxford University Press.

Enghel F (2016) Understanding the donor-driven practice of development communication: From media engagement to a politics of mediation. *Global Media Journal: Canadian Edition* 9(1): 5-21.

Feek W and Morry C (2009) *Fitting the glass slipper! Institutionalising communication for development within the UN*. Report written for the 11th UN Inter-agency round table on communication for development. Washington DC: UNDP and World Bank.

Figueroa ME, Kincaid DL, Rani M and Lewis G (2002). *Communication for social change: An integrated model for measuring the process and its outcomes*. John Hopkins University Center for Communication Programs.

Gagliardone I, Kalemera A, Kogen L, Nalwoga L, Stremlau N and Wairagala W (2015) In search of local knowledge on ICTs in Africa. *Stability: International Journal of Security and Development* 4(1).

Gillwald A (2010) The poverty of ICT policy, research, and practice in Africa. *Information Technologies and International Development* 6: 79–88.

Glennie J and Sumner A (2014) The $138.5 billion question: When does foreign aid work (and when doesn't it)? Policy paper no. 49. Washington DC: Center for Global Development.

Gumucio-Dagron A (2009) Playing with fire: Power, participation, and communication for development. *Development in Practice* 19(4-5): 453-465.

Hornik R (1988) *Development communication: Information, agriculture, and nutrition in the third world*. New York: Longman.

Huesca R (2003) Participatory approaches to communication for development. In: Mody B (ed) *International and Development Communication: A 21st-Century Perspective*. Thousand Oaks, CA: Sage, pp. 209–226.

Inagaki N (2007) *Communicating the impact of communication for development: Recent trends in empirical research*. World Bank Working Papers. Washington DC: The World Bank.

Kogen L, Arsenault A, Gagliardone I and Buttenheim A (2012) *Evaluating media and communication in development: Does evidence matter?* Report written for the Center for Global Communication Studies. Philadelphia: Center for Global Communication Studies.

Lennie J and Tacchi J (2013) *Evaluating Communication for Development*. Oxon and New York: Routledge.

Lennie J and Tacchi, J (2014) Bridging the divide between upward accountability and learning-based approaches to development evaluation: Strategies for an enabling environment. *Evaluation Journal of Australasia* 14(1): 12.

Lennie J and Tacchi, J (2015) Tensions, challenges and issues in evaluating communication for development: Findings from recent research and strategies for sustainable outcomes. *Nordicom Review* 36: 25–39.

Moyo D (2009) *Dead Aid: Why Aid is Not Working and How There is a Better Way for Africa*. New York: Farrar, Straus and Giroux.

OECD Organisation for Economic Cooperation and Development (2008). The Paris declaration on aid effectiveness and the Accra agenda for action. Paris.

Power M (1997) *The audit society: Rituals of verification*. New York, NY: Oxford University Press.

Ravallion M (2009) Should the randomistas rule? *The Economists' Voice* 6(2): 1–5.

Servaes, J. (2016). How "sustainable" is development communication research? *International Communication Gazette* 78(7), 701-710

Shah H (2010) Meta-research of development communication studies, 1997-2006. *Glocal Times* 15: 1-21.

Stern E, Stame N, Mayne J, Forss K, Davies R and Befani B (2012) Broadening the range of designs and methods for impact evaluations. Working paper no. 38. London: Department for International Development.

USAID (2011) *Evaluation: Learning from experience*. Washington DC: U.S. Agency for International Development.

USAID (2016) *Strengthening evidence-based development: Five years of better evaluation practice at USAID, 2011-2016*. Washington DC: U.S. Agency for International Development.

USAID (n.d.) "Meta-analyses, evidence summits, and systematic reviews." Retrieved from http://usaidprojectstarter.org/content/meta-analyses-evidence-summits-and-systematic-reviews.

Vähämäki, J, Schmidt M and Molander J (2011) *Review: Results based management in development cooperation*. Report written for Riksbankens Jubileumsfond. Stockholm: Riksbankens Jubileumsfond.

Waisbord S (2015) Three challenges for communication and global social change. *Communication Theory* 25: 144-165.