

Josh Kaufman: My name's Josh Kaufman, I'm the Director for Evaluation and Impact Assessment in the US Global Development Lab. I thought what we would do is just quickly have the panelists, starting with Michael, introduce themselves. I'll take a minute or two to frame the discussion. We'll have a conversation, but we really want to turn it over, a big chunk of time, to also ... To hear from you, your perspectives, and your questions. So again, Josh Kaufman from the Global Development Lab.

[00:00:30]

Michael: Good afternoon everyone, my name is Michael Woolcock. I work in the Development Research Group of the World Bank here in Washington. I spent most of the last two years based in Malaysia helping the World Bank set up a new office over there. I also have an alter ego life for most of the last 12 years teaching at the Kennedy School of Government.

Buddy Shah: Hi everyone, my name is Buddy Shah. I'm the CEO and one of the co-founders at ID Insight, and we work with governments, foundations, and non- profits around the world to help them generate evidence within their practical constraints to drive improved decision making.

[00:01:00]

Patrick: Hi everyone, I'm Patrick Sommerville from Link. We're an organization that's based here in Washington that's focused on strengthening local [acters 00:01:19] and systems. We work in organizational development, systems analysis, as well as resource mobilization and collective impact.

[00:01:30]

Sarah: Hi everyone, I'm Sarah Bieber. I am a foreign service officer here at US Aid, and I'm currently the Director or our scaling off-grid energy grant challenge for development.

Josh Kaufman: Excellent. So I think I'm still working without that one. Alright. We'll get it right by the end, I promise.

[00:02:00]

[00:02:30]

Thank you everyone for joining us. I hope the morning sessions you've been to have been really good. I've so far been really impressed by the quality of the discussions, so I hope we'll be up to that challenge. I want to start by saying, although I admittedly am biased, I think it's a very exciting time to be a monitoring evaluation research and learning professional. And I think maybe more importantly, it should be a really exciting time to be someone that benefits from our work whether it's folks that are funders of development programs through donor or private foundations, whether it's governments, social entrepreneurs, NGOs. I think we're really seeing the beginning of a blossoming of new types of data and evidence that should be used in new ways. And that's really what we're talking about today, there's the same spirit of experimentation, and risk, and rigor that we're hearing about in the innovation space is very much alive and well, I believe, in the monitoring evaluation research and learning space.

So I'll invite panelists and others to touch on that broadly, but I think in particular, we want to focus on one aspect of that, which is how could we, whether it's

[00:03:00] through new technologies, or new approaches, or even old technologies and approaches that we're using in new ways. How can we generate rigorous evidence and data quickly and timely enough that they could actually be used to inform decision making within programs themselves, right? So we all know there's a lot of great reasons to do evaluation, and research, and certainly monitoring, and we could touch on those, but I think really we want to focus on that particular question.

[00:03:30] And to kick that off I'll turn over ... To start with Michael, with sort of a big picture question, which is ... So I'd sort of referenced ... Oh. We do have, with the advent of broader internet connectivity, and the mobile phones, satellite imagery, there's a whole host of technologies that are increasingly being used by monitoring evaluation research [inaudible 00:03:50] professionals. By projects, by governments, and certainly, like I said, new tools and approaches that are being tested out. How excited should we be about this? Are we at a point where we're beginning to actually overcome the technical and methodological barriers to allow us to feed high quality information into project level decision making? Or is it still mostly hype at this point?

Michael: I guess my answer would be I'm a one and a half [inaudible 00:04:17] kind of guy for a lot of this stuff in the sense that we all know there's plenty of low hanging fruit in developments. And to the extent that some big technical innovations, or even small technical innovations can help us move closer to a possibility frontier where we're able to collect more and better data at scale, and put it back into decision making processes in real time, all of that's a good thing. And as a research guy, more data is better and all that sort of stuff.

[00:04:30]

[00:05:00] But why I'm only one and a half cheers for that because that I think what we see manifest in things like the sustainable development goals is a hyperspace of ambition regarding what we're extensively committing ourselves to do over the next 13 years or so. And precisely because of that hyperspace we have all of these big things that we are trying to achieve for which we can seduce ourselves pretty quickly if we think that we're going to big data our way to the right answer to thinking about these things. If you look at goals number 16 and 17, 16 in particular, "we're going to provide justice for all", right, Really? We've been wrestling with that as humanity for, I don't know, 5,000 years maybe, and we're not ... We're on a very long, slow non-linear path to anything vaguely resembling something like that. And to just kid ourselves that we're just ... That the binding constraint on that not being realized is the absence of a big enough data set to help us do it, then I think we become part of the problem and not part of the solution.

[00:05:30]

[00:06:00] The stuff that I am most excited about is people willing to tackle those kinds of problems, these big questions around justice, these real questions about making school systems work for most people most of the time, making health systems work for people is that people themselves are the binding constraint. Making the human component of those systems function more effectively is what we always struggle with, what humans have always struggled with. And so with the right framing, and the right degrees of modesty, yeah, there's this kind of data can help

[00:06:30] us to move a little bit more in that kind of direction. But if that's what the corkheads end up doing, and we don't have enough people willing to keep wrestling with these really hard problems, then we're just going to end up sooner of later in a world where we have lots of things ... Lots of activity, lots of things being done, but still at the end game with not a lot to show for it.

[00:07:00] And if you look at ... On education for example, since the World Bank's just released its world development report yesterday on education. The stories we see from around the world are pretty depressing actually with regards to the declining quality of schooling over time. Not just that, then huge variations in what governments are able to do with policies notionally committed to achieving everybody to get through primary education. Is the reason that isn't happening because we haven't got data? I don't think the data itself would suggest that's true. I think the data would suggest that people don't show up for work everyday, and sticking cameras in classrooms might be able to help tweak things at the margins on a few of those things. There has to be a whole transformation in the way the human components of organizations are done. And so on a good day, yes, all these big innovations that we're seeing displayed here and elsewhere can be a part of that solution. But if that's sort of where our energy, and our efforts, and our brains get channeled, then I worry that we won't be committing the same amount of energy and effort to these really wrenching problems for which technology is only a ... I would suggest a very small part actually of what the solution is. So one and a half cheers.

[00:07:30]

[00:08:00] Josh Kaufman: Alright. Great, I'll take one and a half cheers. That's still progress I think. So I'd like to frame a question now for Buddy and for Patrick because you both are running enterprises that if everything goes according to plan, gives governments, donors, and other key actors the ability to access high quality information in a form that's suitable for them to then make, if not real time, because we don't live in a real time world, but realer time decisions and changes. What would be your favorite example of that working in the real world? Either something that you've been involved with or that you've seen. Or alternatively, what was a time that it really crashed and burned?

[00:08:30]

Buddy Shah: I can start off. And I think there are a lot of organizations doing great work in this area, I'll just speak about two examples of ID Insights since I know them better. The first is just around some work we did with the Ministry of Health in Zambia where there are very high maternal mortality rates, and they're trying to figure out ways to incentivize mothers to deliver in facilities rather than at home. And basically tweaking a randomized control trial to make it really fast, and give them the evidence they needed before making investments the next fiscal year.

[00:09:00]

[00:09:30] And so we did a randomized trial testing the effectiveness of non-monetary incentives to see if that would increase the number of women that delivered babies in facilities, drawing on global literature and then applying it in this particular context. And essentially we were able to run a really rapid but rigorous randomized trial within six months that fed information into the Ministry of Health, which

allowed them to essentially put these non-monetary incentives into their list of essential medicines for the next fiscal year.

[00:10:00] And so I raised that just as one example to show that the idea of real time or realer time decision making doesn't necessarily mean throwing out the most rigorous tool that we have, and that they can be very flexible if applied in the right way. So that's just one brief example. But I think the one that is a little bit more enlightening for this conversation of what does it take to actually have realer time decision making is some work we did with the Government of Bihar in India and their Ministry of Energy. So Bihar is a state of over 100 million people in India, one of it's poorest and the least electrified, and their binding constraint to actually increasing the amount of electricity rural households got was revenue generation. The households that were getting electricity weren't paying and therefore they weren't able to invest in more generation in transmission.

[00:10:30] And so we had a team that was embedded within the Ministry of Energy that did a whole host of data and evidence activities to first identify what are the sources for low revenue collection. Then second, designed an intervention to actually have private contractors collect the revenues rather than government employees and ran again a rapid randomized trial to see will that increase revenue collection. And we were able to do that, get the evidence, and it found that we did actually increase revenue collection by a substantial margin, but there was still a long way to go. And so first the government acted on that evidence, but then that study threw up two other questions, which is that they increased revenue merely by getting people who had previously paid to pay more frequently. But they didn't actually target any of the people that weren't paying, and getting it for free.

[00:11:00] And so because we were embedded within the government we were able to come up with two or three new ideas and feed back into this process of diagnosing what's wrong, designing interventions, testing them with the most appropriate tool, acting on the evidence, and continually repeating the process. And I think that that's a useful example, at least for me, because it highlights three characteristics that I think are at least necessary, if not sufficient, to have realer time decision making.

[00:11:30] The first is around institutional or political innovations, so I think the idea of moving research [inaudible 00:11:47] not as kind of an external add on to an organization, but part and parcel the decision making process, actually sitting with the government or NGO is extremely important. So thinking hard about what's the political economy of decision making in this place? And how do we structure the evidence generation in a way that fits hand in glove with that political economy of decision making?

[00:12:00] The second is around methodological innovation, I think that having a broad set of analytic tools that you can use for the right question at the right time is really important.

And then the third, which I think is where we're furthest away from the frontier is if

[00:12:30] we actually want more real time decision making there is some technological innovation required to actually get steady streams of high quality representative actionable data at relatively low cost. And in this case we're able to do that because the government already is collecting revenue on revenue data, how many consumers do they have of the electrical grid, how frequently are they paying. And that allowed us to just have much faster iterations and test, and I think that's a big challenge for social programs, is can we build that same monitoring system where we get high frequency data on test scores, on health outcomes, and things like that. But ultimately, I think if you get that combination of institutional innovation, methodological flexibility, and technological progress in terms of the speed of data collection you can do a lot more in terms of real time decision making.

[00:13:00]

[00:13:30] But I would say, to Michael's point, that I'm fairly optimistic, and I think that if we push these three prongs we can have a lot of progress. But ultimately these are still marginal improvements within what is politically feasible in a place. And so if we're thinking about this as creating revolution in how public [inaudible 00:13:28] happens, I think there we have to think much more about the politics rather than technocracy. And there's a role for real time data, but it's probably got to be something that's kind of facing the average citizen and feeding back real time, like political polling information to the people in power, an empowering more activist type stuff. But I do think there's a lot of room for this kind of technocratic innovation.

[00:14:00] Patrick: Great, thanks. Just in terms of my comments. I think I'd start out by noting that we're all working in very complex social and behavioral systems here, right? And that in itself presents challenges, there's no one tool or approach that is going to predict the future for us, or give us perfect information. I think some of the considerations that we need to be mindful of around that are certainly related to the frameworks we're working in, the project frameworks we're working in, the institutional frameworks we're working in, and how enabling that is to learning and adaptation. We also need to be mindful of the context that we're in, and knowledgeable of the context that we're working in, and able to access multiple sources of information.

[00:15:00]

[00:15:30] And then I think, as Michael I think was eluding to, there's a high burden on the people designing and implementing these programs. And being able to drop on resource, identify issues, react, and adapt to them. So I think those for us have been probably the most consistent obstacles that pop up. Just in terms of what we observed, I think maybe just telling a story of how our organization evolved.

[00:16:00] We very much started coming at development work from an organizational development lens, and those of you that know organizational development know that there's a lot of focus on score cards, and capacity measurements, and assessing fixed functional qualities of organizations themselves. We at some point just began asking ourselves why is this not a good predictor of success? Why does this score card not really revealing ... Why is it not really revealing to us who the organizations are and aren't that become successful?

[00:16:30]

[00:17:00] So we really started playing with different approaches and quickly found ourselves in the system's analytical world, and experimenting quite a lot with really two tools; systems mapping approaches, and more actor focused social network analytical approaches. And as that shaped up, I can tell you that over the course of the subsequent years it is main lined into everything we do with local organizations. It's

[00:17:30] there when we advise organizations on strategies, it's there when we facilitate collective impact initiatives. And it's there when we're asked, "How do we design this program?" "Who are the actors we're interested in, and why are we interested in them?" And, "What's driving the system?"

[00:18:00] So I think just judging from certainly where we were in that area of practice five or so years ago to where we are now, we've seen a much greater integration. I think that we can have this conversation around specific tools and approaches, and what's good in one case is probably not good in another case for a long time, but it

[00:18:30] really, I think, again leads back to the facility of the program managers and their ability to assess the situation then draw in tools because none of these can really be taken in isolation. I think the big challenge going forward is going to be certainly on the accessibility in addition to that capacity issue. The accessibility of these

[00:19:00] various tools that we spend a lot of time talking about and experimenting with, and how can we merge the very practical world in which most of us in this room probably reside in terms of hard development challenges in very rural areas, and

[00:19:30] throughout the world. And with the bias on the part of the system's thinker to get it all perfect. So I think there's a lot of work on that going forward.

Josh Kaufman: Great, you can pass it to Sarah if you want. Alright, so one of the themes that I

[00:20:00] heard, which I guess in some respects was a firming of my world view as a recovering democracy human rights and governance officer is that there's a whole different set of constraints that are in some respects independent of the quality, or timeliness of the data. And that really gets at the political economy and the incentives of organizations that have to make decisions based on that, and whether the right incentives exist for them to make those decisions.

So Sarah, you had said I was going to ask a really hard question, and it's true. But I won't ask you to speak for every government or institution out there.

Sarah: Thanks.

[00:20:30]

Josh Kaufman: But at least ... Which would be impossible. But do the nearly impossible and try to speak for USA ID, because we have our own institutional incentives that we have to deal with that often times make it challenging for us to incorporate this sort of information in a timely manner.

[00:21:00] So you are actually, as the Director of our Scaling Off-grid Energy Grant Challenge, you are a consumer of one of these types of innovations in that you have a developmental evaluation evaluator embedded in your project with the hope that it'll help you be ... Make informed decisions and being adaptive. So tell us a little bit

about that thought process, like how did you get yourself into that? And what have you learned so far that might be of broader applicability?

Sarah: Yeah, sure. That's a big question. So I'll probably say something that might be a little controversial in a room full of people that are monitoring and evaluation in research and learning people, but I think it needs to be said that sometimes we have to be ... And this is what I think you, and Buddy were both eluding to, we have to be willing to sort of break the framework a little. So, whether that means breaking your results framework a little, or breaking your scope of work a little, or breaking the program cycle a little to the extent that we can be flexible. We should because sometimes, even though the best plan was laid, there's going to be a political [inaudible 00:21:53] economic situation that we weren't planning for, and I know that was supposed to be captured in our assumptions if we built a good framework. But the reality is the system operates within context, and context in my experience has always driven our decision making much more than the data about what intervention we have done, and whether or not it has been effective. It's almost always a political or contextual issue that drives any shift in any program that we have, for better or for worse.

[00:21:30]

[00:22:00]

[00:22:30]

And so I think to the extent that we can have that data, great. But to make it as important as those other factors is difficult when the framework we're operating in has been set in something that was five years ago, and may no longer be relevant. So I've had the benefit of working with a development evaluator whose been embedded in my team for the past year, or a little less now. And it's completely changed the way that my program has taken shape, and certainly changed the way that I think about how I could be making decisions.

[00:23:00]

[00:23:30]

So I was used to working with a program design, and we have objectives, and we have targets, and we're trying to meet them. And we're doing everything that we can to mitigate any changes to that framework. This person has come in and said, "Sure, sure, sure, everything that you're doing is serving all of those goals and targets. Just do this one thing a little bit differently right now, tomorrow. Go get on the phone and call that person that's going to be making a decision about something that's going to impact your world and make sure they understand the information that we all just learned in this other meeting. Or make" ... And I'm kind of like, "Oh, well tomorrow I'm busy." It's like, "No, this is high priority, this is important", and it's something that I maybe wasn't seeing because I'm heads down in my program.

[00:24:00]

And so I think that has been an enormous luxury to have. I've never had the opportunity to work with somebody that's full time, just looking at whether or not what we're doing is effective, and how to make it more effective. I don't think that's a reality that every single program manager at USA would ever have the ability to have an embedded person full time, all the time. But boy, I really wish it were a reality because it's been tremendously helpful.

Josh Kaufman: I'll just pass it on to somebody. So let me ask another question that's sort of a theme that I heard through at least several of what you were saying, of examples

[00:24:30]

you were using, particularly some of the successful ones. Yeah, one of the trends that I think is helping enable, and it's probably a cycle, but enable this type of reflection between implementation and information is that in a lot of respects the line between implementation and evaluation and learning is getting blurrier. Whether that's where you ... You know, the example Buddy you gave, where it's ...

[00:25:00]

You've got someone embedded in [inaudible 00:25:00] electrical utility in India, or some of these large cross national studies like the J-Pow [inaudible 00:25:07] study where it's like, which is the card and, which is the horse in terms of implementation and learning, and evaluation?

So I personally think that's a good trend, but I'm wondering if you all feel like that's a good trend. And if you think that's something that should be more common than it is now. And anyone can just grab a mic and speak to that.

Buddy Shah:

[inaudible 00:25:30]

[00:25:30]

Josh Kaufman:

Sure.

Buddy Shah:

Yeah. I think in general it's a very good trend. And to Sarah's point, that it's rare to have a full time embedded evaluation measurement expert, I think that that's something that probably should change in the extent to which this is not seen as a nice to have, but a need to have for continual improvement is a broader political shift that probably has to happen for programs to keep getting better. And so in general I think there's just a mindset shift where MNE should be seen as a critical part, and senior role that reports directly to the CEO, or head of an organization, and I'm not sure that we're there yet.

[00:26:00]

And the only downside really that I see of blurring those lines is when you have that kind of embedded relationship you're forced to function within the political constraints of the decision makers. And I think that's really, really positive in terms of making sure that what you're testing actually is actionable and can lead to scale up. But there is a downside to that, which is that you might be doing less frontier, kind of break the mold type of work. And so as there's more blending of implementation and research, I think we should keep pushing that but also carve out enough space for people to be doing work that's kind of outside the bounds of what's politically or administratively feasible at scale now, because we need those early demonstration cases to display, like, "This is what an entirely different way of structuring a particular program or policy could look like". And so I just wouldn't want that to cannibalize that kind of more free thinking approach that a lot of research has done so far.

[00:26:30]

[00:27:00]

Josh Kaufman:

Anyone else want to chime in on it.

Michael:

Yeah. I think there's ... [inaudible 00:27:15] One of the hopeful aspects of all this is a recognition, I think, amongst the development research community, and the evaluators as well, is a recognition that we're not just trying to ask whether questions, like whether this program worked, and up-down, thumbs up-down kind

[00:27:30]

[00:28:00] of sense, but how, and for whom did it work? Those become often the political questions around these things. So when we're talking about health care reform, or we're talking about transitional cash transfers, the more complicated the [inaudible 00:27:44], the higher the standard deviation's likely to be in the outcome space, right? And thus we can construct a mathematical average to say, "Yes it worked", or, "No it didn't". But actually, for these very front line micro level services that are being provided, even or especially at scale. Once you've got big standard deviations you've got to be able to track that, you've got to be able to explain politically and empirically why and how those particular outcomes are being achieved for those particular people.

[00:28:30] So I think, in some sense, those questions are going to ... They're not going to go away, and the only way to answer them seriously is to have a pretty good read on what's going on, and more data on observables is going to help you in that space but a lot of the reason for having embedded people in your research team is because they can actually observe the unobservable [inaudible 00:28:33] these political things actually, that are driving why money gets diverted this way rather than that way. You have no way to tell that from our household data set, you've got to have people that can ... Who have the credibility, and the legitimacy, and the durability to be in those spaces long enough to be able to help answer the questions you're actually asking.

[00:29:00] The other big reason I think is that we're ... When you're in complex intervention space that a lot of us are now in you're not on a linear path from here to there. You're base line and your follow up are going to be a multi pathed journey. And it's not clear after [inaudible 00:29:08] a year whether what you're doing is actually achieving what it's claiming to achieve. Does a null result after two years mean that you failed? Or does two years mean you've just got to hang in there for another five until actually you start to see the change happening?

[00:29:30] So we're doing a lot of work now in the development community of course, on empowering marginalized communities, right? Well we've been working for what? Since Mary Wollstonecraft wrote the Vindication of the Rights of Women to try and get women's equality, and half the people in this room know full well that hasn't been achieved. Have we failed because we're 200 years in, and so we should give it up? Or do we got to go on for as long as it fucking takes until it actually happens? You've got to go with the latter, and you do that as a normative commitment irrespective of [inaudible 00:29:48] independently of evidence. But the fact that we are toiling away up on a vertical cliff and we're not actually seeming to see a lot of rigorous evidence that what you're doing is working or not is not the reason you do or don't continue. Which is to say, you've got to have a pretty good theory, a pretty good theory of change, a pretty good bench marking set of expectations built around what you do.

[00:30:00]

And doubling down on having cooler, better, neater looking evidence, and graphs, and stars on your aggression models, and all the rest of it doesn't help answer those kind of questions. And that's what we ultimately have to answer if we're going to be getting into staying speaking sensibly and responsibly to these really,

[00:30:30] really difficult questions that are already on the table and are only going to get more and more intense as crazy things like the sustainable government goals, as I said, come due in 13 years time when we've got to have indicators on whether we have achieved justice for all, right? Some hapless person in this room is probably going to be charged with coming up with the data set to show that yes indeed, we've achieved this, or something. "Have you? I've probably not, but you've got [inaudible 00:30:51] at least plausibly defend the flat line that you've ... As a noble service in the cause of good because jeez, we're on a thousand year journey here and we're going maybe get there never. But whatever, right? We shouldn't apologize, I don't think, for the very deep normative commitment that mobilized most of us to be in development in the first place. And that mobilized a lot of people to get up every day and do what they do. And evidence is a small component on a good day, to helping nudge that needle in a slightly more hopeful and helpful direction.

Josh Kaufman: [00:31:30] So let me ask a follow up question. It's really based on what you just said, but I think open to anybody, please, which is ... So all 16 sustainable development goals, or sort of any way you want to slice it. A development sector, either the way USA does it or other folks. So clearly they're deeply normative, and some less tangibly normative than others, like justice for all, and inclusion, equality. They're probably not all equally configured to be successful consumers of the types of tools that we're talking about now. So are there particular sectors, or types of interventions that you think for the purposes of sort of deploying resources that are suppose to generate rapid data and evidence, and build in iteration into implementation? Are there particular areas that we should be focusing on, or conversely particular areas we should be avoiding because it's probably not going to ... Even if you ... Either you can't generate the evidence, or it won't get used. And I open it up to anybody.

Buddy Shah: You want to go for it Michael?

Michael: [inaudible 00:32:43].

[00:33:00] Patrick: That's a tough one. I think there are ... Certainly my background, being in the systems approach, world of systems approach [inaudible 00:33:07]. And if we're speaking to, lets say, where we are in the program cycle, design, monitoring, evaluation, I think there's certainly been a number of promising indications that [00:33:30] I've seen in program design, and that functionality in terms of basically empowering the program analyst, or designer, or a manager to have a better understanding of what they're walking into and calibrate their interventions correctly, or more appropriately. I think I've seen a lot of really great developments on that front in the past several years. I think when we start nosing into the [00:34:00] monitoring and evaluation space it gets a bit murkier. There's a lot of different types of programs, and there's a lot of different sources of information that need to be captured. And then again we come back to this burden on the program manager, and correctly deploying tools and approaches where they may be warranted.

[00:34:30] I think that that's where I'd leave my comments on it, but ...

Buddy Shah: Yeah, I could respond.

[00:35:00] So, I mean I think it's certainly correct that we should not have what is measurable drive what is important. And so like that shouldn't set the agenda, and as Michael said, there are normative reasons to pursue justice even if it's harder to gage whether we're making progress against it. However, on the flip side I do think that just because we can't measure something extremely well doesn't mean that some of these same perspectives or outlook can be used to drive something forward. So take sustainable development goals for instance, I think a lot of them we can create real time data systems and evaluation measurement to not just track progress, but also test new ideas and continually improve. And while we might not be able to do that as robustly in the sense of identifiable counterfactuals and statistical significance with justice, I do think that orientation towards articulating what is the goal, finding proxy variables that at least give us an indicative sense of are we moving in the right direction. And then using that feedback to continually tweak what we're doing is something that's widely applicable even when you can't necessarily run a rigorous control trial.

[00:35:30] And so I think that ... You know, I see this move as much more than about methodological innovation, and much more about an orientation towards defining what a goal is, and constantly getting some form of feedback as to how well you're doing towards that.

[00:36:00] Josh Kaufman: Did you want to? Did you want to speak [inaudible 00:36:04].

Michael: Yeah. I can site from a nerdy researcher standpoint, you always start with a question then you work backwards from that to figure out what the tool is rather than have a hammer and find all your nails in the world. So that's what ... I can say that because I'm a researcher. For people who've got to make really hard decisions about budgets, and for people whose job it is to evaluate programs you've got to survive, you've got to keep your show on the road. And so I think there's a ...

[00:36:30] Where one has to be entrepreneurial in a disciplined way with how one goes about this. We can't escape the accountability, and entirely legitimately the accountability requirements we've ... Are all mostly using public money, or money given to us by somebody else to do something. And so there has to be a reckoning at some point with all of that.

[00:37:00] The challenge for me is not so much on the evidentiary side per se. It's what kind of theory, for want of a better word, you use to help interpret whatever the evidence is that you've got. So evidence is never self evident as they just gave before with non-results for example, right? Does a non-result mean that you've got a failure, or does a non-result mean you haven't tried hard enough, or does a non-result mean that it's actually really great and people are working really hard but there's just some big political roadblock stopping it from happening? How do you interpret that? So the results are never self evidently interpreted no matter what the

[00:37:30]

problem is, and no matter what the type of challenges you're dealing with.

[00:38:00] But if we find ourselves in a world where apples, and oranges, and bananas have all suddenly got to be compared with each other, which we often do, then the very strong temptation if you're managing a lot of money under a lot of political pressure is to channel it in the ways that it is the most easily defensible and in a narrow methodological sense. And I think part of the job of research broadly defined is to help protect the space for the claiming that needs to be made in its own terms legitimately around these particular challenges that don't fall as neatly, or as obviously into an evidentiary space where we can wow you with a great presentation that convinces the putative policy maker that this is the best use of their finite resource pool.

[00:38:30] The part of the challenge of being a development professional is protecting the space for a whole array of different things to be done. And not privileging method over analytics, and seeing them as being constantly in conversation with each other. And what I get excited about in development is sort of seeing those moments emerge when reality either forces it just because you [inaudible 00:38:44] groups demand and insist that their claims be put higher up the order. Or just the people who've been doing things in a more narrow way for a long enough period realize that you have to actually explain how and why these things are occurring in the way they are. That there are so many multiple paths by which, especially non-results can be achievable, margin over marginal results, you just need ... With a more, and better theory and method we can throw out these the more sophisticated and accurate our interpretations are likely to be.

[00:39:00] But then the more credible we ultimately are, I think, with [inaudible 00:39:17] ... when we do have our reckoning moment with being asked to explain how this sum of money's been used. That you not just try to shoe horn it into someone else's framing and analytic space because it looks cool, and it's interesting, and it's clean, [inaudible 00:39:33], and it's easy to understand. The hard work of doing this work is being able to protect the space for those really hard things that don't fall into that space.

[00:39:30] Josh Kaufman: Excellent. So I'd like to ... We've got about 20 minutes left, so I'd like to open it up for Q and A. Is this the only portable mic that we have? Okay. I could ... Okay. Might just be a little easier than passing it back and forth. So if anyone.

[00:40:00] Speaker 7: Check, Check.

Josh Kaufman: Back there? So ... And if you could just identify yourself before you ask your question that'd be great. Thank you.

Speaker 8: Well thank you. My name is Estelle Mendez and I work for ACVO, it's an organization that provides both software and advisory services around [inaudible 00:40:23]. But my question has to do ... So my role within ACVO has to do with advising some of the users of our software on good [inaudible 00:40:32] practices. And one of the things that it has come across ... So we work with organizations that

[00:40:30]

are usually implementers, and so this is a question for the funders on the panel. And to what extent are the World Bank, or USAid ready to break the results framework and be open to that more iterative, or there's more, you mentioned developmental evaluation, or yeah, adaptive learning overall?

[00:41:00]

Sarah: I will answer that question in this room.

Josh Kaufman: It's not like you're being filmed or anything, so don't worry.

Sarah: I am being filmed.

I mean, I'd like to say we're ready. I think there are ... One of the benefits of working in the global development lab is we definitely have a team of people that are ready for that kind of change. I don't know that I would say the entire agency is ready to make that move, and I don't think that the systems in terms of the budget are in place to do it. I think that's going to be, always. It always comes down to where the money is, and are we programming our budgets to incorporate that level of adaptability into our programs? And is there space for change? And it just is not how we program our money. I'm not saying it's impossible to do it, and I'm all in favor of it but it's a tough ... I mean, Josh would be the person to talk to about it, but it seems like it's a tough case to make when you have diminishing funds to say, "Well, but a quarter of it we're going to use as open space to adapt and change our program in case things don't go how we thought they would". But I'll hand it over to Josh.

[00:41:30]

[00:42:00]

Patrick: Please.

Yeah. Also, I'm not a donor but I think the topic of ... Maybe there's too much blame in that direction in general. That the blow up the log frame discussion I think is an important one, but we can't underestimate the amount of resistance there is to it within our ranks and among the program managers out in the field. And a huge disconnect that happens between meetings, technical meetings like this and people that need to get work done in the field. And people that need to bound their results, and be able to control their results irrespective of what happens in that environment over the course of the next five years. So I think that's an important dynamic to keep in mind that is often underestimated in these forums.

[00:42:30]

[00:43:00]

Josh Kaufman: So I'll say a couple things. I didn't know if you wanted to chime in. So I'll say a few things. So one is, well I'd like to think of myself as a revolutionary, I'm probably more of an incrementalist. And so ...

[00:43:30]

Buddy Shah: And evolutionist.

Josh Kaufman: Yes, I'm an evolutionary because I've survived 21 years at USAid so I could not be a revolutionary but I actually think that the fact that you've asked that question is hopeful, right? Because one of the things that the lab is a member of an informal group of funders of innovation, and we ... It's almost like a group therapy session

[00:44:00] when we get together because that's one of the key questions in that ... When we say we want to build more of a culture of innovation in USAid, or whether its Global Affairs Canada, or DFID, it's not so much the individual tools, although that would be great, more grand challenges and all that sort of stuff. You know, lets get more mobile technology out there. But it's really about the elements of being able to take risk, and be entrepreneurial. I think in my opinion, USAid is way more entrepreneurial, and willing to take risk than it was when I started. It's not anywhere near where I would like it, but I still ... I feel more confident than not about that question.

[00:44:30]

But I will also say as somebody that plays an evaluator on TV, but I'm not an evaluator, I'm a Democracy Human Rights and Governance Officer, that it also requires a culture change of the researchers and the evaluators because I still run up against the opposite question, which is, "Wait, I can't really answer that question in your world, you practitioner because it's going to make me have to sort of bend my rigid definition of what is sort of like a purely academic study". Which I know as a practitioner won't be useful for me. It might be very publishable and quote right, but I'll never use it. And so I think there's also that world view sometimes that needs to change.

[00:45:00]

Michael: At the World Bank we're trying to explore, in a niche sort of way, not whole sale but in a niche sort of way what [inaudible 00:45:23] called agile development. But sort of just really trying to give a group of people, about 10 each year, the license to try and relieve, and reframe, and rework some of the ways in which we address these complex implementation challenges.

[00:45:30]

In a book I just did with Lan Pritchard and Mat Andrews a few months ago we have a whole chapter on something we call a search frame, which is [inaudible 00:45:43] our explicit attempt to try and articulate what something that should be seen as a compliment to, not a substitute for [inaudible 00:45:50] because the DNA of the bank, and for most big agencies like USAid was technocracy and logistics. It was sort of help ministries of finance and pave a bunch of roads, right? And in that space, that's exactly what a log frame's suppose to do, and 70 years in, it's kind of done okay. But in the 21st century we just find a whole bunch of problems for which we can't shoe horn those problems, and the management of them into this thing called a log frame. It's [inaudible 00:46:16] dumping it, it's about complimenting it with something that's fit for purpose. And nobody's figured that out yet. T

[00:46:00]

[00:46:30] I's a collective task for people in this room, and the development professionals more broadly to figure out how we ensure our accountability norms. How we ensure that we can track the flow of resourcing within an organization when we're trying to do this stuff that is experimental. And experimental by definition means if it's not going to work that means there's going to be this thing called failure. And nobody in this room, or anyone else sort of likes to be in charge of something that is deemed not to have worked. Or is then renders you liable to be hauled before some board to explain why you're not being a profligate steward of publicly entrusted money.

[00:47:00] So there's a whole bunch of implications that go with actually trying to allow people to do this thing called experimentation and adaptive development. And there's a whole raft of administrative clauses that have to be in themselves experimented with if we're going to be able to do this work at scale. So we're still, as this was indicated, the fact that we're even asking those questions itself is a [inaudible 00:47:22], one hopes, of a broader social movement amongst development professionals that's opening up. But it's because there isn't an answer to that particular problem, we've all got to figure it out collectively. But doing so, I think, ultimately with a spirit of this is going to ultimately be a compliment to, not a substitute for the orthodoxy.

[00:48:00] Orthodoxy, I've been around long enough to know that orthodoxy [inaudible 00:47:44] is there for on a good day for a good reason, and it helps us do the kind of stuff that was originally set up to do. Why keep coming back to the hyper space of the sustainable development goals, they were not on the agenda in the 1950's when our organizations were conceived, now we've got this vastly different array of things we need to achieve, and vastly different means of trying to pull them off. And just trying to shoe horn all of that into something designed to make the defense the department work is just not in anybody's world the right way to do it. But we should [inaudible 00:48:16] the same sort of creativity and diligence that our grandparents did, and try and figure out the solution to this problem. And groups like Buddy's and others on the panel are part of that process, I hope all of you are as well.

[00:48:30]
Josh Kaufman: Over there. And then we've got our next question will be over there.

Speaker 9: Hi. Nida with IRI, quick question. Are we doing enough evaluation of secondary impacts, beneficiaries of our beneficiaries? Is there potential catalytic data there that we should be mining? And what about old school ways of evaluating, are we losing sight of the individual story, just like spot checks? I feel like with a lot of the systems, and a lot of the SNA we're missing individual stories, I don't know if that's true or not.

[00:49:00]
Buddy Shah: Can you just elaborate what you mean by beneficiaries of beneficiaries?

Speaker 9: So, we tend to evaluate primarily the direct beneficiaries of our programs, particularly amongst the implementers. But I'd like to ... I haven't seen a lot of work around evaluating secondary impacts, so their beneficiaries to understand whether our work is truly impactful.

Buddy Shah:
[00:49:30] Yeah. So I mean I don't have a broad enough view of whether people are measuring spill over effects and externalities enough. I do think that there is work that can be done [inaudible 00:49:34] to say, "Do we think based on theory that this program is going to have spill over effects, or effects on those proximate beneficiaries, and then invest the resources to actually measure them". And so I think that it probably

just has to be a case by case basis based on theory of whether it makes sense to invest those extra resources in measuring those spill over effects. So yeah, that's what I'd say about the first portion. And what was the second question?

[00:50:00]

Speaker 9:

Sorry. Just on whether or not we are still gathering individual stories, and if we're missing those datas going back to sort of orthodox ways of thinking about monitoring evaluation.

Buddy Shah:

Yeah. So I mean, my view on that at least is I think stories and narrative are really powerful for shifting big political things, but I don't think the pendulum has swung so far towards the side of data over stories that I'm particularly concerned about it right now.

[00:50:30]

Patrick:

I think too. I think you observed that maybe the user experience, or the narrative is getting lost in a lot of the emphasis on data and systems, which can be a bit technocratic and impersonal. There are approaches that focus more on narratives, perspectives, user experience of beneficiaries of constituents in development programs that seek to ... And really, I think at least in the systems world, the hallmarks of those would be approaches that are very deliberate about [inaudible 00:51:26] and capture of diverse voices, and broad listening. And explicit about where that inquiry is narrowed. And also, a bit iterative in terms of coming to answer, and coding, and categorizations of themes, and things like that. So there are, I think, those kind of less impersonal approaches out there that are being actively used.

[00:51:00]

[00:51:30]

[00:52:00]

Josh Kaufman:

We have a question right over there.

Speaker 10:

Hi my name is [inaudible 00:52:08], and I want to talk about an experience I had and ask you to comment about the potential of [inaudible 00:52:14] systems to also engage political officials, and engage communities. A year ago I was working in the [inaudible 00:52:20] of Peru and we implemented [inaudible 00:52:23] system for schools, it was centralized. But we basically collected information about different indicators, one of them was attendance of students, and principals, and teachers. And one of the results we noticed is that by establishing this cycle of giving the information back to local governments about how they were performing, we changed there mind set of starting to look at how was their performance, and also compare it to other local governments neighbor to them. And also [inaudible 00:52:52] potential of making this information available to the public, and also engaging communities to make them more accountable. So I would like to know if you can comment about that possibility of also in monitoring systems. Thank you.

[00:52:30]

[00:53:00]

Josh Kaufman:

Buddy, I know a lot of your folks are embedded in government so do you want to speak to that?

Buddy Shah:

Yeah. I mean, so just ... You know, we're doing a similar project with the central

[00:53:30] government of India. Essentially rating different districts on some composites, so socio-economic index. And then getting those district bureaucrats to compete with one another based on progress towards them. And so I think that that's hugely valuable, and it actually ... Speaking of pendulum, I think one thing that may have been overlooked in this move towards more data is basic monitoring systems. And so I think that evaluation rightfully has gained a lot of momentum but perhaps at the expense of just simple monitoring systems that are meant to track, and then incentivize competition or progress. And so I personally think that I would love to see a lot more investment in that space, even as the evaluation space grows.

[00:54:00] Michael: I think that's a really interesting observation because I think ... I'm a sociologist, so I can say this, you're trying to change a culture around evaluation not just the methodology around evaluation. So in all the stuff that Michael Kramer for example, and all the debates they had around the de worming, and all that sort of stuff. I think for me, all the fights that researchers want to have over was that study properly identified, or did they get their counterfactuals right?

[00:54:30] When you hear them actually talk about what they did in Kenya, right? The real impact to me is that they've changed how public officials talk to each other, and how they talk to beneficiaries, how they talk to citizens about why they do and how they do what they do. And only researchers care about whether their model was specified the right way, right? But the bigger impact I think, really is about trying to normalize a public discussion around evidence around some minimal threshold of quality as a [inaudible 00:54:52], or as a prelude to a broader and more sophisticated conversation.

[00:55:00] Now we all in this room, I imagine, that's Nervana, right? [inaudible 00:55:00] where everybody geek out on the quality of the data that they're using to have a broad policy dialogue on these kinds of things. But I don't think we should underestimate the importance of that. I think that we wanted to, perhaps a self interest in wanting it to be more narrowly technical because then it privileges our particular education and the salaries we're getting paid to be the consultant in that space. But the really big challenge happens when you do a lot of the work that my friend, and colleague [inaudible 00:55:29] is doing with his social observatory in

[00:55:30] India or other places where you're really helping beneficiaries themselves to become little researcher. You're helping them to be able to be not just literate with regard to data, but not just be people that provide data to these white people that come in and write fancy journal articles, you're helping them themselves to be able to be savvy users, and are insistent on getting good data be part of that conversation.

[00:56:00] So I think that's a really important compliment, and kind of almost inevitable that will happen if you do this at scale. But I think it's a big victory actually when citizens themselves see themselves not just as objects of external researchers coming in to try and write nice papers. But as co-contributors to the production, and the use, and the interpretation of data about themselves. And if that can happen, then I think that's ... I would actually rate that much higher. As the public good it's

[00:56:30] created out of the research process, if that's the outcome then I think we should figure out how to create a metric to really assess that because I think that's a big, big win.

Josh Kaufman: Great. We have time for one, maybe two more questions. Yeah, why don't we take two real quick.

Speaker 11: Hi Dalton Brantly, the Mitchell Group. And we do monitoring evaluation in [inaudible 00:56:53] Africa with USAid primarily. And I've a rather simple question about how do you sell innovation, in a sense, both from a funding prospective and an implementing prospective? Everyone cares about cost but you can frame it from other angles like high quality, you can make it faster, you can do all these different scenarios. So I'd be interested to hear what your strategies are to make an innovation digestible to someone you approach with a certain tool, or innovation?

[00:57:00]

Josh Kaufman: Okay, and is there another question out there?

Speaker 12: You know, the way the conversation's moved in the last two questions and answers does seem to be reflective of the fragmentation in which all these systems operate. And flexibility, and ability to [inaudible 00:57:37] frame is great. At the end of the day it's about harmonization of a lot of development efforts on the ground in different communities. So explicitly as practitioners, to what extent can we overcome that fragmentation and move towards harmonized systems? Is there evidence ... And I know there's a lot of meetings, and certainly in the HIV there's been certainly the one [inaudible 00:58:00] framework, and one authority, and one programmatic framework movement. And we've seen this in a lot ... I'd imagine with the sustainable development goals we'd have the same kind of move to harmonize outcome measurements especially, and work locally to make that happen rather than the unitary developer in the program cycle managing a project. And we've got to move away from the project, so can we explicitly address the need to harmonize not only USG efforts, which are fragmented, but also European donors who [inaudible 00:58:34] money, et cetera. I mean there's a lot of developments about fragmentation, how can we overcome that I guess is the premise, and could be more explicit about that.

[00:58:00]

[00:58:30]

Josh Kaufman: Alright. So I'll take a stab at the first question maybe, and I'll try to do it quickly because we want a couple minutes, I want someone to answer the second question.

[00:59:00]

So I would say in terms of marketing innovation, we're getting people excited about it. I would say two things really quickly. One is relevant to the particular types of innovations we were talking about here, which is innovations in the monitoring evaluation research and learning space because one of the things we're doing in the lab right now is we have a series of innovations that we're testing out with USAid missions and bureaus sort of operating as the pilots that are explicitly monitoring evaluation research and learning innovations. And so there I think it was a matter of ... There's a series of sort of problems, from want of a better term, that I think there was sort of clear agreement on. One is a lot of the traditional

[00:59:30]

[01:00:00] tools were designed for more linear ... The log frame sort of model in predictable environments, and with a clear theory of change. And we all know that that especially increasingly, those aren't the environments that we were working in. So I think on the part of folks in the agency, and the missions, and the project managers there's a real hunger for new approaches that'll actually let them measure more accurately the things they're doing instead of trying to have ... Use the hammer and nail analogy that you used. I think ...

[01:00:30] More broadly, selling inter marketing innovation, that's a really tough one from a USAid perspective because we have so many stakeholders. So the one place where I think we've really hit it out of the park is with the innovators themselves. And that's because there was a lot of innovation going on regardless of whether AID was funding or not, right? Which is one of the most exciting things going on in development, right? So when we set up these tool, grand challenges, development innovation ventures like five or six years ago, we sort of had a sense for what was out there but it was sort of like field of dreams where we built it and there was a huge demand on the part of the people that wanted to solve problems, right?

[01:01:00] I think the greater challenge is how do you demonstrate the results of those innovations, like you said better, cheaper, faster, in a way that your other key stakeholders, whether [inaudible 01:01:00] stakeholders, whether they're people in other federal agencies, or even within USAid, see the value of it. I think that's still an ongoing process. And I think we'll ... Ask me in a year or two and I'll let you know how it went. But I think that the innovators themselves ... I think that that's been a real huge success story, at least from a USAid perspective.

Does anyone want to tackle ... We've got about another minute, that second question. That was a really hard question, I don't want to answer it. So I want someone else to try to do it.

[01:01:30]
Buddy Shah:

[01:02:00] I'll [inaudible 01:01:30] 30 seconds. So I mean, at least from my point of view, I think harmonization has to start with the return of political sovereignty to the nation state in which you're doing work. And so I see huge difference between my experiences working with World Bank in India versus in Zambia or Kenya. And so I don't know the feasibility of this, but I would say I'm less optimistic about a technocratic, bureaucratic harmonization effort between a bunch of multilateral funders, and more optimistic of a situation in which it's very clear that the Minister of Finance or Health in Kenya puts his or her foot down and says, "Look, this is the direction we're going. You can get on board if you want." And that's been the approach of the government of India, and while there's a lot of frustrations working with Indian bureaucrats, I think that it's probably directionally absolutely the right way to go.