



USAID
FROM THE AMERICAN PEOPLE

PROGRAMNET

TRANSCRIPT

COMPOSITE INDICATORS: AN INTRODUCTION TO THEIR DEVELOPMENT AND USE

AUGUST 1, 2019

Presenter(s): Jerome Gallagher (PPL/LER)
Mark Skeith (E&E/PO)

This document was produced for review by the United States Agency for International Development. It was prepared by the USAID LEARN contract. This document is the transcriber's best effort to capture the words spoken by the presenters. It may contain minor transcription errors. For full accuracy, please consult the webinar recording.



[0:00:00]

Jerome Gallagher: All right. Hello and welcome to this webinar on composite indicators, an introduction to their development and use. We are very pleased you are able to join us this morning. I'm Jerome Gallagher. I'm an M&E specialist on the institutional support services contract in the PPL Bureau and one of the things I do in PPL is try to provide guidance, technical assistance on program monitoring. So I'm very excited about our topic today.

Composite indicators is a topic I find both interesting and increasingly important in our work at USAID. I hope you will find that as well and that's why I'm so delighted to be joined today by Mark Skeith. Mark's an empirical analyst from the E&E Bureau. Mark has been instrumental of development of USAID self-reliance metrics and just has load of experience working with composite indicators, reviewing composite indicators, developing composite indicators.

[0:01:00]

I'm convinced he probably even dreams about composite indicators in his sleep. Welcome, Mark.

Mark Skeith: I have, on occasion, dreamed.

Jerome Gallagher: [Laughs].

Mark Skeith: Thanks, Jerome. Great to be here and hi, everyone. Really excited to speak about this important topic. Composite indicators have really ballooned in popularity across development policy circles and even among practitioners in their programs including increasingly within USAID. With their rising popularity, we've seen the practice composite index construction improve dramatically in recent years which has been great to see but we've also seen some persistent critiques. In some cases, well founded ones.

The reality is really that composite indices are good tools when they're purpose built, when they're built responsibly, when they're built transparently and when they're used by users appropriately. And they are bad tools when they are not. So I'm glad to see how many people selected they're fans of composite indicators in the initial poll.



[0:02:00]

We only have one hour today and this is just a primer but we do hope this webinar helps to equip you to interpret the barrier of composite indices out there to better inform your work.

Jerome Gallagher:

All right. So let's just go quickly through the agenda for today's webinar. So first, we're just going to say a few introductory points about composite indicators, what they are, some strengths and limitations. Next, we're going to briefly go through some but not all of the key steps in designing a composite indicator from the underlying conceptual framework, to visualization. This is a big topic so we're not going to get into a lot of details but I think understanding some of these steps will help if you ever need to develop or review or even just be a consumer of a composite indicator.

Then we'll talk about reviewing a composite indicator with a little bit of a focus on trade off and then we'll point you to some resources –

[0:03:00]

where we think are useful for learning more about composite indicators. And then at the end, we'll open it up for questions that you might have about composite indicators that we might be able to answer.

Mark Skeith:

Okay. Thanks Jerome. So let's start with a quick definition of a composite indicator or index. Essentially, it's a set of indicators that reflect various dimensions of some unobserved singular concept. So when taken together, those various measures gauge that singular concept. The index's ingredients and how they are aggregated are determined by the author. It's key to understand what decisions were made by that author and why and the purpose of the index. Today, we're going to walk through some of the building blocks of composite indices and these building blocks are all highly interrelated in most cases.

One of the key points to take away that I hope you take away here today is that every composite index that has ever been built is built on some form of tradeoffs and judgment calls.

[0:04:00]



This is unavoidable and not a reason to disregard composite indices all together. It's no different than any evaluations that's ever been conducted or survey that's ever been administered. The key is really whether the index is built reasonably, transparently, and based on evidence when possible. One quick example, it's probably one of the most widely known examples in the development field is the UN's Human Development Index. As this graphic shows, it uses four indicators to gauge three pillars of human development, income to reflect a decent standard of living, life expectancy to reflect a long healthy life, and two indicators of years of school and to reflect the knowledge base. It's a pretty straight forward index but there's a lot that goes into building even something like this.

Jerome Gallagher: So Mark, what are some of the strengths of composite indicators?

Mark Skeith: I wanted to highlight two main strengths that I see. One big advantage is obviously that they reduce the visible size of a set of indicators without dropping the underlying information base.

[0:05:00]

Although, they do, of course, obscure that underlying information base. By doing so, they are easier to interpret than a battery of separate indicators, especially for senior leadership or for non-data folks or for people outside the development field. So essentially, they can provide a quick, accessible bottom line in digestible format. Another big strength of composite indices is that I think it's relevant to USAID is it can fill some key knowledge gaps by allowing for critical comparative analysis of multi-dimensional issues that simply otherwise wouldn't be possible.

So for instance, there are many of these types of multi-dimensional issues that we gravel with every day here at USAID like competitiveness or fragility or disaster risk. They can't be captured by a single indicator and require far too many indicators to be adequately and feasibly summarized for decision makers, policy makers. Jerome, how about now, can you highlight some of the limitation of composite indicators?

Jerome Gallagher: Sure. Yes.

[0:06:00]



For all the composite indicator skeptics out there, important to mention the limitations as well. So first, I think these indicators can take a lot of time and effort to develop. So there really needs to be a compelling case for the added value that a composite indicator brings or we're just having a set of indicators that individually measure the various phenomenon of interests rather than trying to aggregate them into a single indicator.

Second, by summarizing concepts that are multidimensional nature and aggregating multiple indicators into a single indicator, these composite indicators can disguise the variation across cases and invite some simplistic and often inappropriate policy conclusions. For example, we often try to measure democracy. Right? This big concept of democracy.

[0:07:00]

But two countries that score – that have the same score on a democracy index, right, might have very different reasons if they score low, for instance, for why they are considered undemocratic. Right? One might be more about the competitiveness of elections in that country while in another country that receive the same score, it might be more about lack of freedom of the press. So an aggregate measure can disguise some of those variations if you don't look at – if you don't disaggregate it into the individual indicators and that can invite simplistic policy conclusions and policy responses.

And then just a last reason, design and construction of these composite indicators can be quite complex and that can often obscure the inner workings of that composite indicator and really just make it hard to determine what subjective decisions were made in the construction of that composite indicator –

[0:08:00]

or if there was bias in its creation or just poor measurements in some of the indicators that make up that composite indicator. So at this point, you might be asking why we should even care about composite indicators. Well, despite, I think, some of the limitations, composite indicators are something that I feel we have seen a lot of interest at USAID and I think rightly so. Of course, the J2SR country road maps provide what are probably the most prominent composite indicators in USAID right now. Road maps use 17 third party indicators, some of which are indices



themselves and aggregate them into composite indicators of country capacity and country commitment.

But I think they did a pretty good job on that *[laughs]* and I know Mark –

Mark Skeith: Thanks, Jerome.

Jerome Gallagher: Had a hand in that. But in addition to the self-reliance metrics, those are really just the tip of the iceberg. Numerous missions are using composite indicators throughout their program cycle.

[0:09:00]

In many cases, missions are drawing on third party country level composite indicators in their CDCSs or in their performance management plans to help engaging progress toward development objectives. We also see missions working with implementing partners to develop new composite indicators at the project or activity level. For example, the Bosnia mission has, for many years, worked with an implementing partner on the judicial effectiveness index. And I understand Uganda mission is working on composite indicators to measure resilience at the regional level. Even our standard foreign assistance indicators that we use to report in the annual PPR includes some composite indicators.

For instance, the civil society organization sustainable – sustainability index which is one of our DR standard indicators.

[0:10:00]

So with that, let's get into talking about the design or the development of composite indicators. And we'll start with arguably the most important element in the development of composite indicators and that's the conceptual framework. And by that, we just mean that the theoretical basis for the selection and combination of variables or indicators into a meaningful composite indicator. So a clear conceptual framework is really the foundation of constructing the composite indicator. It drives the subsequent steps in the development and construction of a composite indicator.

What is badly defined is going to be badly measured so we really want a good conceptual framework as our foundation. And it just involves defining the phenomenon that you aim to measure,



determining what factors matter for that phenomenon and determining their relative importance.

[0:11:00]

And those determinations can be shaped by theory, by empirical research, political philosophy, the academic literature, et cetera.

And of course, different composite indicators can have different degrees to which they're based on empirical work or particular world views. And in new policy areas, sometimes, there's not a lot of work done on a particular phenomenon, which is all the more reason then, to invite experts and stakeholders into thinking through the theory underlying the conceptual framework underlying a composite indicator and being part of its construction.

Mark Skeith:

Great. Thanks, Jerome. Let's take a look at an example now of a couple conceptual frameworks and how their organization can help drive results even if they're trying to measure the same thing.

[0:12:00]

In this case, we're looking at two different composite indices that are both aiming to gauge economic freedom worldwide. At left, we have Frazier institutes index of economic freedom and at right, we have Heritage's index of economic freedom. Like most indices, these frameworks both take the form of a nested or tiered structure. When you look across them at the content, at first glance, they look very similar covering many of the same topical areas like government size and the legal system which isn't exactly surprising because they both aim to gauge the exact same concept, economic freedom and are using all the same data.

But when you look closely at the structure and the contents of each index, you can see a few major differences that can influence the results substantially. One quick example is when you look at where trade freedom shows up in their framework, it features much more prominently in Frazier's version than in that of Heritage. It's one of Frazier's five main pillars which it is only one component of heritage's market openness pillar.

[0:13:00]



As a result, the variable ends up constituting 20 percent of Frazier's overall score, yet only 8 percent of heritage's score. Same can be said of monetary health, as well, if you look. And there's plenty of other examples if you dig into this and really, when you compare any two indices that are trying to gauge the same thing. There probably are important differences in their structures that drive the results.

Jerome Gallagher: So Mark, why are there pumpkins on this slide?

Mark Skeith: [Laughs] That's a good question, Jerome. So for those listening here and apologies in advance to the non-Americans who are going to be thoroughly confused by this analogy but I'd ask everyone out there, what criteria do you use when you go to a pumpkin patch to pick a pumpkin? Is one criteria color? Maybe size, integrity of stem? Maybe whether the pumpkin has weird bumps all over it. Of course, there's no absolute best type of pumpkin. It really depends what you're using it for and your preferences. You may simply prioritize seed volume if you're baking seeds.

[0:14:00]

Or you may even want those warts if you're trying to paint it into an ugly, evil witch. So with that analogy, I'm not going to read all of the criteria you use when selecting indicators for an index but as with pumpkins, there are a lot of different criteria to consider when selecting indicators that feed into an index and really, there's no right or wrong answer as to which criteria are going to be most important for you and for your indexes purposes. But I will say that the relative importance is inherently tied to the index's intended use and also, the nature of what you're trying to measure, the concept you're trying to measure for that indicator.

But I will say typically, two central factors tend to be the relevance of the indicator to the phenomena being measured and obviously, the reliability and accuracy of measurement. So if we're reviewing a global index across countries, we really want the indicator to ideally, be a reliable predictor of the concept being measured across as many countries as possible and over time.

[0:15:00]

And when data are scarce, you may need to look for proxy variables and to determine whether it's better to include those



verses omitting the concept altogether from the framework. So it's a good example of one of the tradeoffs that need to be made by an index constructor.

So let's take an example from our agency's recent development of the new journey to self-reliance country road maps where we were tasked with finding the 17 best suited indicators for measuring what were identified through the agency through a huge consultative process that has 17 core interrelated aspects of commitment and capacity in each country. Hopefully, most of you are familiar with this roadmap tool but real quick, the 17 metrics organized into seven sub-dimensions, which in turn, are grouped into two dimensions, capacity and commitment.

So let's take a look specifically at one of the seven sub-dimensions government capacity, which comprises three indicators shown here.

[0:16:00]

One of those indicators is efficiency of tax administration, which is a very important aspect of financing self-reliance and an emerging priority in the development field at large but in our case, for this process, it ended up proving to be a tricky one to pin down in terms of finding a single indicator that is meaningful, reliable, and timely across all developing countries or even a majority of them. So at right, we can see some of the main candidates we considered applying the criteria that I showed you on the pumpkin slide. We were look for something that could either capture tax administration or even better, the broader concept of domestic resource mobilization.

So the first candidate, government taxed revenue as a percentage GDP, simple captures total tax revenue collected as a percentage GDP. And it's beneficial in that it's very easy to understand what's being measured. It's based on objective concrete data and covering is pretty good. However, we found that relevance across countries is questionable because the optimal level that you'd want to see would really depend on the country's situation and economy.

[0:17:00]

So because of that comparability issue, we shelved that option.



Another candidate we looked at was the World Bank's country policy institutional assessment measure of efficiency of revenue mobilization. This one really hit the head – nail on the head conceptually in terms of the key issues we cared about but coverage was really poor. Essentially, a non-starter for the roadmap. And the quality of the methodology is a bit of a black box, which is another ding in the con column so we moved on.

Another candidate, World Bank's paying taxes was a part of the – is a part of the well-known World Bank doing business index. For this one timeliness, coverage across countries were strong but it's very narrow measurement wise in that it's centered only on the experience of a typical firm in the largest city in each country with no consideration of other firms or even of individual income taxes or property taxes. Things like that.

Also, it did not correlate even loosely with any of the other candidates which was a huge red flag for us so we moved onto the next candidate.

[0:18:00]

We also considered an indicator called tax effort from USAID's collecting taxes database. This indicator resolved the key weakness with the government tax revenue candidate in that it measures how much tax revenue a country collects vis-a-vis what can be reasonably expected of that country to collect given its macroeconomic, demographic and institutional features. However, it is part of the USAID brain and data set and for at least for the roadmap's first year, we were advised to stick with third party data to remain objective for that initiative.

So we ended up settling on the institutional profile databases efficiency of tax administration indicator. It nailed exactly what we wanted to measure conceptually. It had decent coverage. It was just updated in 2016 and it was third party. It's a qualitative assessment with seemingly credible results for the most part. The problem with it really, one of the big problems with it is it's only updated every three to four years so it was a bit of a temporary solution.

[0:19:00]

We actually reevaluated these options again this year for the FY20 roadmaps, in part based on mission feedback and priorities for



this year. This year's roadmaps. And we're going to actually revise this metric again for the sake of continual betterment of that tool. I'm not going to go into that here but look at or more information on all the change to the roadmaps in a couple webinars PPL's hosting next week.

Jerome Gallagher:

Thanks, Mark. I think it's very helpful to go through that process of understanding how one selects the indicators for a composite indicator. Next, we – so now, that you've thought about a conceptual framework and you've selected indicators for a composite indicator, the next element in the development of a composite indicator is normalization and that's just a process of transforming the measured units of each variable or indicator in your composite indicator so that they're all on the same scale.

[0:20:00]

It's quite typical that when you select indicators, they'll have different scales so one might be percentage. One might be accounts. So normalization is a process that one needs to go through in developing that composite indicator. Here's just three of the most common approaches for normalization. There are many, many different types of approaches. The first is just a simple rank, right? So if you're creating a composite indicator of ten countries, each indicator will be transformed into a number from one to ten depending on the country's rank on that indicator.

The next is one of my favorite, min max method. It sets a scale where the lowest value is zero and the highest value is one. We'll go through an example in a minute. And then there's also standard scores. There's Z scores also quite commonly used. It normalizes indicators to a mean of zero and a standard deviation of one. Of course, as I said, there are many ways to normalize indicators, each of which have their strengths and weaknesses.

[0:21:00]

And choices about normalization should be informed by both the properties of the data, such as the presence of outliers and how the data's expected to be used. Whatever choice is made, it really should be documented with the reasons for that choice. Something to look out for when you're reviewing the composite indicator. Let's just go through a really quick example. Here, I've created what I called a regional generosity index and I wanted to create this composite indicator by mashing up two indicators, a



percent of the persons in each region who volunteer their time and the charitable donations per capita in each region. And you can see the regions on the left.

So let's normalize them with the min max method and we're going to focus on the East region in particular. All right. So here's our formula for the min max formula of normalization.

[0:22:00]

Pretty simple. If we look at the East region and our volunteer indicator, we just take the value for East. We subtract the minimum value, which is North at 5, and then we divide it by the max minus the min. So 32 minus 5 and we get a score of .33 for the east region. All right? And then I've done the same thing for all of the regions and normalized the volunteer indicator. And you can see we get scores between 0 and 1.

You could do the same thing for the charitable donation score. So in the East, we had \$119.00 per capita. Take that, we subtract the minimum, which was 53 and then divide it by the max minus the min and we get a score in the East of .15.

[0:23:00]

And then I can do that for all of the regions. Now, I have both of my indicators normalized so that they have a minimum of zero and a max of one. And then, we're going to get into which is the next subject, weighting in aggregation. I'll just quickly do that for this composite indicator. Just simply give them equal weights. Right? Add them up. Divide them. Get a mean score. So the mean score for East, the index score for East is .24. And then I can do the same for all the rest and I now have a composite indicator for the regional generosity index.

But I was getting a little bit ahead of myself in talking about weighting aggregation, so let me turn it over to Mark to talk about weighting in aggregation.

Mark Skeith:

Thanks, Jerome. So weighting can be equally, if not more, influential on the final results than any other step in the framework.

[0:24:00]



Perhaps the conceptual framework and weighting are equally as important building blocks. Generally, they're less correlated and index's ingredients are the more influential the weighting becomes. How much weight an author places on each variable can be based on the sources organization's policy imperative or development theory or the author could crowd source the decision, tapping perspectives of subject matter experts to assign weights. There are also statistical techniques that can be applied to drive weighting. There's not enough time to get into those here but you can check the resources at the end of the webinar to learn more about those are just reach out to use if you have questions.

Generally, those apply more complexity to an index even – so while improving precision, they may be decreasing accessibility in terms of users being able to understand them. One big thing I wanted to stress on the topic of weighting was that equal weighting is still a weighting decision. It's still a judgment call to say that all things matter equally for all countries worldwide.

[0:25:00]

Of course, that may not be the case. As particular variables might be important for your country and others not, some sources even allow you to assign weights on your own to the components of an index so that you can build a picture that's tailored to your context or to the decision you're trying to make. And now that interactive dashboards are easy to create, it's easy to provide this feature to users so that they can explore the index's inner workings more deeply for this reason and for many others. It's typically best to use equal weights when the conceptual model is not based on strong, evidence based.

In these cases, the index's purpose is typically, more to bring attention to an emerging issue or to spark debate over dynamics as it is for the purpose of precise measurements. So that has been the case with the journey to self-reliance country road maps. That's how those have been handled. Essentially, those are USAID defining its world view for what matters for –

[0:26:00]

what generally matters for commitment and capacity worldwide and thus would have been foolish for us to do anything more than equal weighting at this point. But we're very transparent about



that and very transparent with this being really to center strategic thinking around those concepts and to garner reactions for what matters in countries for the sake of continual learning. And that tool will probably continue to evolve over time.

One other point on the concept of aggregation, one quick point on that one is to only aggregate up to a level that's meaningful and no further. Again, using the country road maps as an example, we don't want to go – we didn't want to go to a sort of meta self-reliance index because that wouldn't have been particularly meaningful and we wanted to center people's thinking on commitment and capacity. So you really want to go to a point that's appropriate to what you're measuring and no further. So let's take a look at a quick example using GSMA's very interesting, useful, mobile connectivity index.

[0:27:00]

On this one, I wanted to bring this example to your attention to demonstrate the interaction between weighting in aggregation within the conceptual framework and how the sort of interaction between the two collectively drives how much indicators influence overall index results. So here, you can see the mobile connectivity index's four pillars. Each of those have equal weighting at 25 percent. Let's look at two of those - sorry. There we go. Let's look at two of those sub-pillars that content and services and infrastructure and the indicators that go into them.

Yeah, sorry. I'm having trouble with my slides. So if you want to know how much a given variable is influencing overall results, you need to look at the weighting scheme, the organization of the conceptual framework, and how all of those ingredients of the index are aggregated upward to reach the overall score.

[0:28:00]

Here, I want to compare two indicators that are highlighted in red. The first is number of mobile apps available in the national language and the second is international internet bandwidth per internet user. When you – to understand how much the two indicators are influencing overall results, you need to take into account not only their weights which are circled here in blue, but also the weights of their respective sub pillars and pillars, also circled in blue. Taking all of that into account, you can see that mobile apps indicator accounts for five percent of the overall



mobile connectivity index where the internet bandwidth indicator accounts for only 1.5 percent of the overall results, once you run all the math and do all the aggregation.

So given the conceptual framework of this index and the weighting that they've applied and the resulting aggregations, the indicator in red atop has more than three times the effective weight on the overall index results for an indicator at bottom even though they look equally as prominent at quick glance.

Jerome Gallagher: All right.

[0:29:00]

So we've gone through conceptual framework, selecting indicators, weight – normalization, weighting, aggregation. Last element that we want to talk about is visualization.

Mark Skeith:

Thank you, Jerome. I'm glad – really wanted to cover this as well because it's a topic that's often overlooked but when you talk about garnering attention to your index or driving how users interpret results, visualization is very, very important building block in that composite index. It's a very important step. The manner in which data are presented can encourage various types of comparison points. They can help bring out the story underlying the data. You can use visualization as a tool to steer users in how they interpret the results and can make the results more accessible and obviously, if you have compelling visuals, you can draw interest to your data.

We could do a whole webinar on this topic in and of itself but generally, we want visualizations to be clean, clear, and aesthetically pleasing.

[0:30:00]

You want to ensure users can interpret what is being depicted while minimizing so called noise and clutter in its presentation. You want to remove anything that isn't necessary and you can always drop caveats and disclaimers to the bottom because if people are really interested in getting into the weeds, they will get into the weeds. This visual at right shows some of the many visualization ideas you can find at a very useful resource called Data Vis project. It's a website linked at the end of the webinar. If done well, eye popping visuals can be advantageous to captivate an



audience and provide insights to users but I also would caution you can also go too far offering something that looks fantastic visually but really provides no analytical value and may confuse people that aren't data experts.

So let's look at an example here. Going back to our Frazier versus Heritage index of economic freedom example, this slide is showing their overall results as they present in their most recent reports. Let's see how each organization depicts its overall index results visually and how that influences how users might interpret the results.

[0:31:00]

Again, as we saw before with the framework, conceptual framework, at first glance, these two resources look quite similar but there are some key differences in the presentation that will influence user interpretation. First and most obvious, Frazier at left is encouraging comparison of your country of interest to the rest of the world while Heritage is encouraging regional comparisons. Second, the two sources bucket countries into categories differently.

Frazier is using a math based approach, splitting countries into four quartile groupings whereas Heritage splits countries into five groups based on seemingly arbitrary score cut offs that they feel are appropriate, which, of course, is fine, if they made those decisions in a reasonable manner. But this decision can lead to fairly different interpretation of the results. Looking specifically to East and Central African, you can see a fairly different picture between the two sources. Kenya looks more like Tanzania in Frazier's version but more like Ethiopia is Heritage's version. You also lose a lot of variation because of the colors that they've chosen and the groupings.

[0:32:00]

You lose a lot of variation across much of Central Africa in Frazier's version, whereas in Heritage's version, it really highlights the Chad, Sudan, Central African Republic, and Republic of Congo are particularly concerning. And these differences in how the results are being interpreted might has as much to do or more to do with the visualization decisions as it does actually meaningful differences in the results. So you really have to dig into the data to determine precisely what's going on here and when you're



making a visual, being conscious of how users may convey and garner takeaways from your visual.

Even something as seemingly minor as using white book country border as heritage does helps differentiate between countries and improve readability whereas Frazier does not. All right. We're going to skip the second vis for the sake of time here so we can save room for Q&A. I'll hand it over to Jerome now to show you a really, really useful resource that as you're assessing an index and its components.

Jerome Gallagher:

Yeah. So thanks, Mark. We've talked about the different elements of developing a composite indicator.

[0:33:00]

And most of us in our jobs at USAID won't actually be creating composite indicators but I hope that understanding some of these elements of composite indicators can make us better consumers of composite indicators. And many of us will need to review composite indicators perhaps as CORs of an implementing partner who creates a composite indicator or as MEL specialists who are looking for third party indicators to help us in monitoring our country context. So to that end, we developed a little checklist of sorts to help with reviewing a composite indicator.

It's in the file share box that you'll see at the end of our presentation. It's about – it's mostly about just identifying the topic that you want to think about and consider when reviewing a composite indicator. And of course, one of the main topics that the checklist touches on that we haven't really talked about is tradeoffs.

[0:34:00]

Developers of composite indicators need to balance different tradeoffs. So it's something for reviewers to think about as well. So let's just go through a couple of the tradeoffs mentioned in that checklist.

Mark Skeith:

Thanks, Jerome. As I mentioned at the beginning of the presentation, it's very important for every index to be designed for its purpose and audience and as such, there's no right or wrong way to construct an index. This fit for purpose principle should guide each decision throughout construction and these



tradeoffs have great bearing on the internal validity of an index and its policy relevance. And it's really up to the craftsmanship of the author as to whether there's real value and credibility and validity in the resulting data.

Whether the author is aiming to gauge human development worldwide or measuring something like crop field productivity across districts for an activity, there are some common tradeoffs that they face and these tradeoffs are by no means inevitable but they are frequently in play. One key tradeoff is between simplicity and complexity.

[0:35:00]

There's always an inherent tendency toward complexity to capture all that matters, to maximize accuracy relevance and even stakeholder buy in to ensure that everyone's interest and priorities are reflected. Yet, there's also a very real practical need for simplicity so that the index and its ingredients are easily understood and accessible. Top level accessibility is really invaluable for a complex index as most user will probably not explore beyond the top level structure. So a user who can easily understand the overall structure and the quote on quote ingredients of an index will feel oriented and more prepared to buy in.

But the more complicated you make the structure for the sake of analytical rigor, the more difficult it will be to communicate what goes into it. So that's the tradeoff that has to be made there depending on your user base and what it's measuring. One example is the fragile states index and that's from the fund for peace. And to achieve what they feel are credible results, they end up factoring so many data points and conducting a really complex content analysis.

[0:36:00]

And this may simply be necessary to get at the really complex content of fragility but as a result, we can't see the math behind the scores and it's impossible to know what is driving the results and how that content analysis is really conducted when you get into the sort of mechanics of it. Some analysts, as a result, aren't comfortable using what is essentially, a mythological black box. Although, I personally, still think it's a very useful tool and the results have seemed credible with everyone I presented them to.



Another important tradeoff is between continuity and relevance. An index builder's understanding of an issue may evolve over time as I'm sure it will with us with the self-reliance concepts where new valuable data may become available that wasn't previously. This author has been faced with the decision as to whether they want to revamp the index's design to quote-unquote, remain with the times, or stick with their current approach that is familiar to users and allows for historical trend analysis. It's kinda hard to sum up which one to prioritize.

[0:37:00]

Obviously, it's going to depend on each case but I will say that while it has its downside, change generally can add credibility and typically makes for a better index in the long run. One good example is the World Bank's well-known doing business index. Following many years of critique, the bank ended up undertaking some substantial revamping to their methodology over the past five years or so. But now, as a result, comparability across years, especially to result before 2015 is not really advisable anymore.

The last tradeoff I want to talk about here is between coverage and precision. Precision in measurement often comes with substantial LOE as any M&E specialist on the line knows. This is especially true for indicators gauging things like institutional capacity or listed activities. So for a country level index, this often means that some of the more meaningful rigorous measures that could feed into an index do not have strong coverage over time or across countries. The index builders are often faced with a decision between a strong metric covering less countries or time versus a less relevant measure, perhaps even a proxy measure that offers greater coverage across countries than over time.

[0:38:00]

One good example is the World Justice Project's Rule of Law Index. It offers a much more impressive mythology, in my opinion, than the World Bank's world governance indicators for in terms of collecting on the ground data in each country and how it's measuring the issues of rule of law. They draw on substantial public polling in each country, annex for questionnaire data in each country. But as a result, world justice project is only currently able to cover 126 countries, although that's quickly



expanding, while World Bank covers more than 200 countries and has done so for more than two decades now.

Jerome Gallagher:

Thanks, Mark, for that discussion of the tradeoffs. Those are some things I hope everyone will think about as they either develop or review composite indicators. So this is – we're almost at the end of the presentation portion of this webinar and we hope that the short introduction encourages you to learn more about composite indicators.

[0:39:00]

If you download the slide deck after the webinar is over, we have some references listed so you can learn more. One in particular that I'd like to just mention, which is also in our file share box, is the handbook of constructing composite indicators from the OECD. It's just a great resource that goes through all of these topics in much greater depth. Some of it gets technical but even if you're not mathematically inclined, it still just includes a bounty of really useful and accessible information to dig into.

And also, another way of just learning more about composite indicators, one of the best things you can do is just dig into composite indicator that you're interested in that's fairly established or well-known.

[0:40:00]

Look through the methodology section in the documentation of that indicator and then try to search out some critiques of that composite indicator to see what some of the choices are that were made – some of the choices that were made in developing it and how others might have made different choices. It can really be an informative way of understanding composite indicators more.