

What cannot be measured still must be managed.^{1, 2}

Dumas introduced new characters, like the wonderfully taciturn valet, Grimaud, who never replied with anything more than a monosyllable. Actually, this was an ingenious trick on Dumas's part, for newspapers paid by the line. Rapid-fire dialogue had the double advantage of making the serial easy to read and of multiplying the author's revenue tenfold.

This was all spoiled the day *La Presse* and *Le Siècle* decided no longer to consider a line any word group that failed to extend at least halfway across a column.

That afternoon Villemessant, editor in chief of the *Figaro*, happened to be at Dumas's. He watched the author reread a manuscript and strike out entire pages.

"What are you doing, Dumas?"

"Killing him."

"Whom?"

"Grimaud. . . . I made him up for short lines. Now he's worthless."

From "Alexandre Dumas: A great life in brief" by André Maurois.

With scarce resources to support development and poverty reduction, it is unsurprising that the demand for "results" has increased, and with it more demand for the quantification of the results of development projects. How much benefit do we get from projects financed by tax-payer money? What are the best projects that deliver the most benefits? What are the sectors that contribute the most to development and poverty reduction? What are the best organizations that spend their resources most efficiently?

This demand is particularly pressing regarding "resilience", as this topic has attracted more interest and resources in the past few years. Building resilience is now an explicit objective in multiple strategies and projects, and large amounts have been spent with this goal.³ Metrics that focus on "inputs"⁴ show an increase: more projects aim at increasing resilience; we have developed new tools, methodologies, and databases; we are investing more in adaptation and resilience-building;⁵ and there are more people working in the field around this endeavor. But what about the results or outcomes of these efforts? How

¹ Stephane Hallegatte and Nathan Engle, with contributions, edits, and comments from Marianne Fay, Benoit Lefevre, Julie Rozenberg, and Sundus Siddiqi, May 2018.

² This work is a product of the staff of The World Bank. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. Nothing herein shall constitute or be considered to be a limitation upon or waiver of the privileges and immunities of The World Bank, all of which are specifically reserved.

³ Williams, A. (2016). Options for resilience results monitoring and evaluation for resilience-building operations. World Bank Group & GFDRR Scoping Paper.

⁴ Organization for Economic Co-operation and Development - Development Assistance Committee's (OECD-DAC). (1991). Principles for the evaluation of development assistance.

⁵ EBRD, AFDB, ADB, EIB, IDBG, and WBG. (2016). Joint report on multilateral development banks' climate finance; Buchner, B. K., Oliver, P., Wang, X., Carswell, C., Meattle, C., & Mazza, F. (2017). Global landscape of climate finance 2017. Climate Policy Initiative.

much resilience has been “produced” with these resources?⁶ It is of course a reasonable and important question to ask, if resources are expected to increase in this particular area in the future.

It is not only a question for the monitoring and evaluation (M&E) of the use of public resources, but also for the design of development plans and projects. As the saying goes, “what does not get measured cannot be managed.” Many feel that we, as development practitioners, will become better able to increase resilience – and through increased resilience, improvements to peoples’ well-being – if we can measure it in a quantified way. Good indicators for the resilience generated by projects would help select the best projects to build resilience, and create appropriate incentives for teams and institutions to ensure that their activities contribute most effectively and efficiently to this objective.

Resilience related indicators are also sought by the private sector, for example to identify the riskiness of investing in a certain project or set of projects, or defining a preferred asset class that contains more “resilient” investments or more “resilience-building” investments.⁷ There is interest in such metrics from social-impact investors as well, and good resilience metrics would facilitate the mobilization of private capital toward resilience-building projects.

However, the quest for a resilience indicator has been difficult, and existing indicators are not fully consistent. While many indicators have been produced using different methodologies, there is discord on a single metric. Indicators like ND-Gain or INFO-RM aggregate many sub-indicators – from the number of people exposed to risk to the presence of a risk management institution in a country – to generate simple metrics for countries’ vulnerability to climate change or other crises. The World Bank’s “socio-economic resilience” indicator uses modeling to assess the ability of a country’s population to cope with and recover from disaster losses. These indicators provide different views on various aspects of resilience, and can help identify priorities or guide policy action; however, they do not provide a unique consensus indicator that can be used to evaluate a portfolio of projects or track progress over time.

One challenge is the existence of multiple definitions and scopes for resilience and risk. A resilience indicator can only measure the resilience *to something*, such as rapid-onset natural hazards like hurricanes, climate change and other long-term trends like desertification, or man-made shocks like conflicts or civil unrest. Different indicators include different events in their definition of resilience, which explains part of the difference across their estimates. But even for the same set of events, definitions can vary. For instance, some include the exposure to natural hazards in the definition of resilience (a country affected by regular typhoons will be less resilient than a country where typhoons are rare, everything else being equal). The World Bank’s socio-economic resilience indicator – published in the *Unbreakable* report⁸ – instead considers resilience as a complement to exposure (the number of typhoons hitting a country does not affect its socio-economic resilience, which only measures the ability

⁶ Organization for Economic Co-operation and Development. (2014). Guidelines for resilience systems analysis: How to analyse risk and build a roadmap to resilience; Bours D., McGinn, C., & Pringle, P. (2014). Selecting indicators for climate change adaptation programming: Guidance for M&E of climate change interventions series. Guidance Note 2. SEA Change, UKCIP.

⁷ Koh, J., Mazzacurati, E., Trabacchi, C. (2017). An investor guide to physical climate risk & resilience: An introduction. Global Adaptation & Resilience Investment Working Group (GARI).

⁸ Hallegatte, S., A. Vogt-Schilb, M. Bangalore & J. Rozenberg. *Unbreakable. Building the Resilience of the Poor in the Face of Natural Disasters*. The World Bank. Washington DC.

of the population to cope and recover *when* a typhoon hits and causes damages). In the *Unbreakable* framework, the overarching measure for the impact of natural disasters on a country is the *risk to well-being*, which combines the exposure (how often the population is affected), vulnerability (how much damage is caused by natural hazards when they hit the country), and socio-economic resilience (how capable the population is to cope with and recover from the damages caused by hazards, when they hit).

Here, we use the term “resilience” in its broadest sense, and include shocks (such as natural disasters like storms) and stresses (long-term trends like climate change and desertification). An action to build resilience can act by reducing exposure (e.g., making sure the population is not affected by building outside of flood zones, reducing the share of climate-sensitive sectors); by reducing vulnerability (e.g., reducing the damages when the population is affected by a flood with appropriate building norms, or making agriculture climate-smart); or by building socioeconomic resilience (e.g., improving access to insurance and improving the quality of reconstruction, or improving the ability of workers to shift to sectors that are not affected by climate change).

In addition to indicators that measure the resilience of a country or community, other initiatives have analyzed how to best measure the resilience created by a project, looking at different sectors and different components of the project.⁹ At the World Bank, such efforts have concluded that each project is different and can produce resilience through a variety of different channels; thus resilience-building at the project/program level is highly context specific. As a result, although it is possible to develop a theory of change and/or results framework that assesses the resilience related outcomes for a given project, these indicators are highly unlikely to be capable of aggregation into a single number to assess the resilience produced by a portfolio of projects.¹⁰

Still, pressure on the World Bank and other MDBs is increasing to design an approach to systematically measure resilience benefits through a single/aggregated outcome metric. It is therefore important to flag the risks of over-relying on imperfect quantified indicators to measure progress and to prioritize future investments for resilience building. Particularly important is the problem with how the indicator – and its inability to measure what we want – may lead to perverse incentives.

Other sectors where quantified indicators have been used for management purposes can offer some useful insights.

- **Education** professionals often rely on quantitative metrics, for instance to measure school performance. But, they are also aware of the risks when such metrics occupy too much space in decision-making. The use of standardized tests in education has been criticized for (at least) three reasons: teachers’ role arguably extends beyond what can be easily quantified (there are no easy metrics for the ability to reason); tests provide imperfect measures of what they try to

⁹ Note that measuring the resilience created by a project is different from ensuring that a project is resilient. For instance, a new road may be resilient – i.e., able to manage natural hazards and climate change without excessive interruption and repairs – without making the community more resilient. The need to make all new projects resilient is addressed using climate and disaster risk screening (to identify the lack of resilience) and appropriate decision-making processes and project design methodologies (including what is referred to as “decision-making under deep uncertainty”).

¹⁰ World Bank. (2017). Operational guidance for monitoring and evaluation (M&E) in climate and disaster resilience-building operations. World Bank Group.

measure (for instance, they measure at one point in time); and attribution is difficult (a bad score may not be due to a bad teacher or a bad school, for instance if the students are from disadvantaged backgrounds). The perverse incentive here is that many studies suggest that performance-based pay for teachers has led teachers to “teach to the test,” leading to score inflation,¹¹ with little or no demonstrable improvement in education in general.¹²

- In **health care**, the use of quantified indicators has led to similar issues. Bevan and Hood report, for instance, that hospitals in the UK made patients wait in ambulances outside emergency services to ensure that their wait time inside the building would not exceed 4 hours, the threshold used in hospital performance metrics.¹³ Even more problematic, poorly designed indicators can push practitioners to avoid the most difficult cases. Since 1992, annual risk-adjusted mortality rates are public for all hospitals and surgeons providing Coronary Artery Bypass Graft Surgery in Pennsylvania. In a survey realized in 1996, fifty-nine percent of the cardiologists reported increased difficulty in finding surgeons willing to perform surgery in patients with highest mortality risk, and 63 percent of the cardiac surgeons reported that they were less willing to operate on such patients in response to the publication of mortality rates (in spite of the risk adjustment process).¹⁴
- Similarly, imperfect indicators in **police** work and in the **justice** system have led to the wrong incentives and undesired behaviors. In France, there is evidence that some policemen shifted their focus from solving crimes to tracking illegal immigrants, as it was an easier way to increase the number of arrests (one indicator included in the decision-making regarding performance bonuses). In the French judiciary system, indicators and incentives to reduce the duration of cases has led to a successful decrease in the average duration – but a closer investigation shows that only the longest cases have experienced a reduction in duration, while the duration of short cases has been increased. Overall, most cases have seen an increase in duration, not a decrease.¹⁵
- When **unemployment agencies** have been evaluated using the share of placed workers, a very reasonable indicator at first sight, they shifted their efforts on the individuals who are most likely to find a job at the expense of the most vulnerable people.¹⁶

¹¹ Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and management*, 30(3), 418-446.

¹² Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521-546; Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging Growth: How to Judge No Child Left Behind?. *Educational Researcher*, 36(5), 268-278; Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources state NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, 34, 209–231; Reback, R., Rockoff, J., & Schwartz, H. (2014). Under pressure: Job security, resource allocation and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, 6, 207–241.

¹³ Bevan, G., & Hood, C. (2006). What’s measured is what matters: targets and gaming in the English public health care system. *Public administration*, 84(3), 517-538.

¹⁴ Schneider, E. C., & Epstein, A. M. (1996). Influence of cardiac-surgery performance reports on referral practices and access to care—a survey of cardiovascular specialists. *New England Journal of Medicine*, 335(4), 251-256.

¹⁵ Bacache-Beauvallet. (2011).

¹⁶ Anderson, Burkhauser, & Raymond. (1993).

- And of course, the role of bonuses based on short-term performance has led to excess risk-taking in the **financial sector**; a phenomenon now well-documented in the aftermath of the most recent economic crisis.

One common dynamic observed in many of these cases is that when the indicator is unable to properly account for the difficulty of an action (even with risk-based indicators, like in health), professionals tend to focus on what is relatively easier. However, there is often higher value in tackling the most difficult cases – for instance, because the worst crimes are the most difficult to solve or, in our resilience case, because the people who are the most vulnerable to natural hazards are living in low-capacity environments where projects are particularly challenging to implement; and where data and capacity to track progress are often harder to come by.

Another concern is how complementing intrinsic motivation (i.e., the willingness of individuals to do their job properly) with extrinsic motivation (e.g., a monetary reward based on performance) can in fact reduce the intrinsic motivation, to the point where the overall motivation is reduced by the additional monetary incentive.¹⁷ For example, when day cares introduced a fee that parents have to pay when they arrive late to pick up their children, parents tended to arrive later more often.¹⁸ Because they pay when they are late, the moral imperative of being on time appears less important, and the net impact is to reduce the incentive to be on time. The same effect has been observed for the willingness of local communities to accept undesired local projects, such as airports, new chemical or nuclear plants, or prisons.¹⁹ Providing monetary compensation to the people living close to an undesired project can reduce their willingness to accept such a project for moral and ethical reasons (as a contribution to society or their community) and can make people even more reluctant to accept the project than without compensation.

As we develop indicators to measure the resilience benefits of our projects, the lessons from these sectors are important to keep in mind: we should not forget that indicators create incentives, and that bad indicators create bad incentives. While an imperfect indicator could appear “good enough” to track the performance of an institution, it can be dangerous if staff start to focus their effort on “looking good” according to the indicator. Like the teachers who “teach to the test,” development professionals may start to “develop to the indicator,” with potentially negative impacts on peoples’ lives (and reduced efficiency in the use of scarce resources). This effect has been referred to as the Campbell’s law, which states that *“the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”*²⁰

A few “thought experiments” around the risk for an indicator to create bad incentives for development professionals and institutions can inform whether institutionalizing a resilience indicator makes sense. The objective of the exercise below is to better anticipate how the indicator could realistically affect incentives and decision-making, to ensure that it does not lead us toward the wrong projects. Each experiment considers one or several alternatives for a project, with one clearly superior

¹⁷ Bénabou & Tirole. (2003); Falk and Kosfeld. (2006).

¹⁸ Gneezy & Rustichini. (2004).

¹⁹ Frey & Oberholzer-Gee. (1997).

²⁰ Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and program planning*, 2(1), 67-90.

to the other(s), and explores how different indicators can create an incentive for selecting a suboptimal solution. The experiments cover seven characteristics that we want projects to demonstrate: (1) efficient, (2) context specific, (3) fair, (4) transformational; (5) comprehensive; (6) robust; and (7) difficult.

Efficient. Consider a new road that will be built in a low-income country. A risk screening exercise shows that the road will be exposed to large flood risks. Two options are on the table: option A reinforces the road and increases drainage capacity, which makes the road more robust but increases its cost by 30 percent; and option B moves the road by 2 kilometers so that it avoids the flood zone, at a marginal additional cost. Here, the question is whether the indicator will favor the costly solution at the expense of the smart one. Input-based indicators (e.g., “how much of my project finance am I able to attribute to climate change adaptation or resilience?”) risk favoring and incentivizing the more expensive options.

Context specific. Consider two communities in very different situations. Community A suffers from minor droughts because it has no water storage and is completely reliant on rainfall for farming. Farmers in this community cultivate crops that can survive drought, but have low productivity. As a result, poverty is widespread in community A. Community B is richer and uses irrigation massively, applying groundwater pumping to provide the water, which leads to rapid salinization of the water reserves and creates a threat to longer-term water supply for agriculture and human consumption. Here, the question is whether a resilience indicator can capture the nuance that community A can become more resilient due to investments in irrigation and moving to more water-intensive crops, while community B can become more resilient due to dis-investment in irrigation and a move to less water-intensive agriculture. Resilience in these cases would indicate moving in exactly opposite directions in terms of the associated resilience-building activities in each of these communities. Similar issues will arise with adaptive social safety nets, which may either make a community sustainable and resilient by facilitating coping during a bad year, or in other situations may keep people in locations where they have no credible long-term prospects, pathways to prosperity, and decent living conditions. Positive lists of adaptation action (e.g., all irrigation projects count as adaptation to climate change) will struggle to take into account the context of the operation and may lead to favoring one-size-fits-all solutions and maladaptation (when the resilience-building activities are considered across broader spatial scales and longer temporal dimensions).

Fair. Consider two coastal protection projects. Each project costs \$100 million. Project A is expected to prevent \$20 million in flood losses every year, while project B is expected to prevent \$10 million in annual losses. On the basis of a classical cost-benefit analysis, project A appears superior. However, the difference in benefits arise from the fact that project A protects a very wealthy neighborhood while project B helps an informal settlement where people are poor and do not have a lot of assets to protect. While the rich neighborhood could finance its protection itself (but is happy not to if somebody else pays for it), the informal settlement cannot pay for its protection. Furthermore, while project A reduces asset losses from floods without many development benefits, project B helps many people escape poverty over time as they do not have to spend the little they save on repairing their homes on a regular basis. Thus, indicators for resilience of climate change adaptation projects that are based on “avoided losses” or monetary benefits anticipated by the project will favor projects similar to project A (and more generally any projects protecting richer regions and individuals, which concentrate most of the economic value in a country). A resilience indicator thus needs to be subtle enough to include equity and poverty considerations, and needs to look beyond direct benefits to consider the development

dividends from different operations; otherwise, those projects will be incentivized that might not produce the best dividends for the intended beneficiaries.

Transformational. Think of a coastal protection investment project at an early stage of project development. It targets a city where about 100,000 people are flooded every year or every other year. Considering available resources, two options are on the table: focusing investments in the highest risk areas, which cover 10,000 people, and providing them with protection against the 100-year event (i.e. the event with a 1% probability of occurrence every year); or spreading investments over the whole population, which with the same resources reduces slightly the flood frequency, from one in two years to one in three years, but for the whole population of 100,000. Indicators that focus on the number of beneficiaries (e.g., number of people with increased resilience) would of course drive business toward projects with marginal impacts but a lot of beneficiaries. And these projects are the least likely to be transformational to help people escape poverty for good. Instead, indicators should take into account not only the number of beneficiaries, but also the extent to which the project can change their life and generate sustained benefits.

Comprehensive. Think of an urban development project in a large coastal city. A risk screening flags – unsurprisingly – that climate change and disaster risks need to be considered in the design of the project. In response, the team adds a stand-alone component to the project: financing coastal protection. Five years after the project is completed, people realize that the protection system is insufficient and that the project has inadvertently exposed 100,000 people to flood risk, leading to massive retrofitting costs. Here, the danger is that adding one component may give the impression that the problem is solved. And indeed, the “simple action bias” is a well-documented behavioral bias:²¹ when confronted with a risk or a problem, we tend to do one thing and consider the problem solved. Similarly, a team working on an urban project may decide to add a couple of parks to absorb more rainfall, and will assume that all flood related risks have been addressed. But, disaster and climate risks are sometimes so high that they cannot be simply managed by a separate component but need to be accounted for in the design of the full project. When using an indicator for resilience in development projects and plans, teams must ensure that it will not push the client toward singular actions or box-ticking, but will incentivize more significant resilience-building interventions. This logic also cascades up to the need, oftentimes, to account for the risks beyond the project level. That is, the risks and resilience-building measures developed on a project-by-project basis might not be adequately accounting for the relationship with other activities occurring within the country, nor the implications of the resilience-building measures beyond the individual project cycle. Instead, a desirable resilience indicator would need to be sufficiently situated in country planning and broader programmatic approaches for investments, and would need to manage the multiplicity of scales, from the national to the project-level.

Robust. Consider a hydropower investment that is highly dependent on water availability and rainfall. The team hires a climate science consultant to provide information about future climate conditions at

²¹ Weber, E. U., and E. J. Johnson. 2011. “Psychology and Behavioral Economics Lessons for the Design of a Green Growth Strategy: White Paper for World Bank Green Growth Knowledge Platform.”; Weber, E. U., P. G. Lindemann, H. Plessner, C. Betsch, and T. Betsch. 2008. “From Intuition to Analysis: Making Decisions with Your Head, Your Heart, or by the Book.”

the location of the dam to inform the project design. The consultant uses one of the leading climate models, which projects an increase by 35 percent of rainfall at the chosen location, and the project design is adjusted accordingly. Since climate change has been taken into account in the project design, many indicators would count the hydropower dam as “resilient” or “climate informed.” But then of course, the consultant may have used a different leading climate model, and the projected change in rainfall could have been a 20-percent decrease, which would have led to a completely different design. Future climate conditions are highly uncertain, and there are regions, like West Africa or India, where models even disagree on whether rainfall will increase or decrease in the future. Some indicators may label a project as resilient that has been designed taking into account future climate conditions, or has used climate model outputs. Such indicators may encourage teams to design their project using a unique climate model to make it “climate informed” at the lowest possible cost, without considering the uncertainty in future climate conditions. A resilience indicator should incentivize projects that are “robust,” i.e., projects that deliver development benefits under a wide range of possible climate and socio-economic conditions.

Difficult (enough). Consider two projects aiming to improve the resilience of agriculture production to drought in two different communities. The first one takes place in a community with high capacity, and its chances of success are close to 100%. The second one would have a major impact, as it targets a very poor and vulnerable community with low capacity. But the latter project faces political and technical obstacles and its chances of success are estimated at only 25%. Indicators that do not account for the level of risk of the intervention could easily create an incentive to focus on safe (but marginal) projects, at the expense of more ambitious and transformational ones (this was observed in the health and police sectors, even though health indicators are usually adjusted for risk). Since we expect that the contributions of international organizations will help countries tackle difficult issues, this bias is clearly undesirable. Note that if we try to correct for this bias by explicitly favoring risky projects, then we will fall in the opposite problem and will favor projects with excessive risk levels. The finance sector provides a telling illustration, as traders have asymmetrical compensation schemes that are bounded at zero in case of large losses, and therefore have an incentive to take too much risk.

Consider one resilience-building project – say the retrofitting of all hospitals to enable them to resist storms and earthquakes – that can be implemented in only one of two regions in a country. Region A has seen a lot of progress in terms of resilience in the last decades, pushed by targeted and efficient action by the government. Region B is fragile, with recent conflicts, and has seen limited action from local authorities in disaster risk reduction. Even an indicator to measure resilience perfectly would incentivize actions in region A at the expense of the arguably more-in-need region B, if the indicator can measure resilience but cannot attribute resilience gain to a specific intervention. Indeed, 5 years after the project is completed, it is likely that resilience in region A will have kept improving, making it possible to make a loose connection with the project and the observed project. Such an indicator would drive action toward places and countries where gains are easier to achieve. This would also incentivize teams to support projects that would have been implemented anyway, i.e., projects that crowd out domestic actions and are not “additional” compared with a no-intervention scenario. To avoid this bias, it is possible to request projects to provide a “theory of change,” i.e., a causal link between the intervention and the gains in resilience, and indicators able to measure the existence and magnitude of this link.

These thought experiments may not summarize all the challenges and drawbacks that a crude indicator creates. But being aware of these risks can help prevent some of the negative consequences of using these indicators.

Some of these risks and problems can be mitigated by *improving the indicator*, and ensuring that it can capture the various dimensions discussed here. For instance, an indicator measuring the number of beneficiaries could also include a threshold of how much gain each beneficiary benefits from. To ensure projects targeting the poorest and most vulnerable people are not at a disadvantage – even though the monetary benefits from these projects may be smaller and their implementation risks larger – individual benefits accruing to the poor and rich can be valued differently, as the World Bank did in the *Unbreakable* report, to measure socioeconomic resilience. Similarly, a process-based indicator measuring whether projects are designed taking into account climate change could require that a range of climate models are used to stress test the project, and be based on strict procedural norms regarding how these different climate model outputs are taken into account to reach a robust decision.

Risks created by imperfect indicators can also be reduced by *combining several indicators*, possibly with additional guidance or rules. If an indicator is based on aggregated benefits, and therefore risks favoring better-off beneficiaries over poorer ones as highlighted in the example above, then a complementary indicator could be the same measure of aggregated benefits, but only counting people that are in the bottom 20 percent in terms of income. While introducing multiple indicators may be seen as complex (and costly in terms of time and resources for operational teams), a set of complementary indicators is less likely to lead to large negative outcomes than a single indicator.

Regardless of the quality of an individual indicator/metric and any associated complementary indicators that can be developed, it is evident that aggregated quantitative resilience metrics will only take us so far. Resilience is as much about infrastructure and financial instruments as it is about governance, voice, and empowerment. But governance, voice, and empowerment are not easy to quantify and measure, and should not be the losers of the overuse of quantified metrics.

In the face of the complexity of the issue, making development more resilience-oriented will require that the tools used for project prioritization, design, and M&E have enough flexibility to include resilience in the most relevant way. This is likely at the expense of an aggregate resilience metric that could be used to measure the resilience generated by a portfolio of projects that are different in nature.

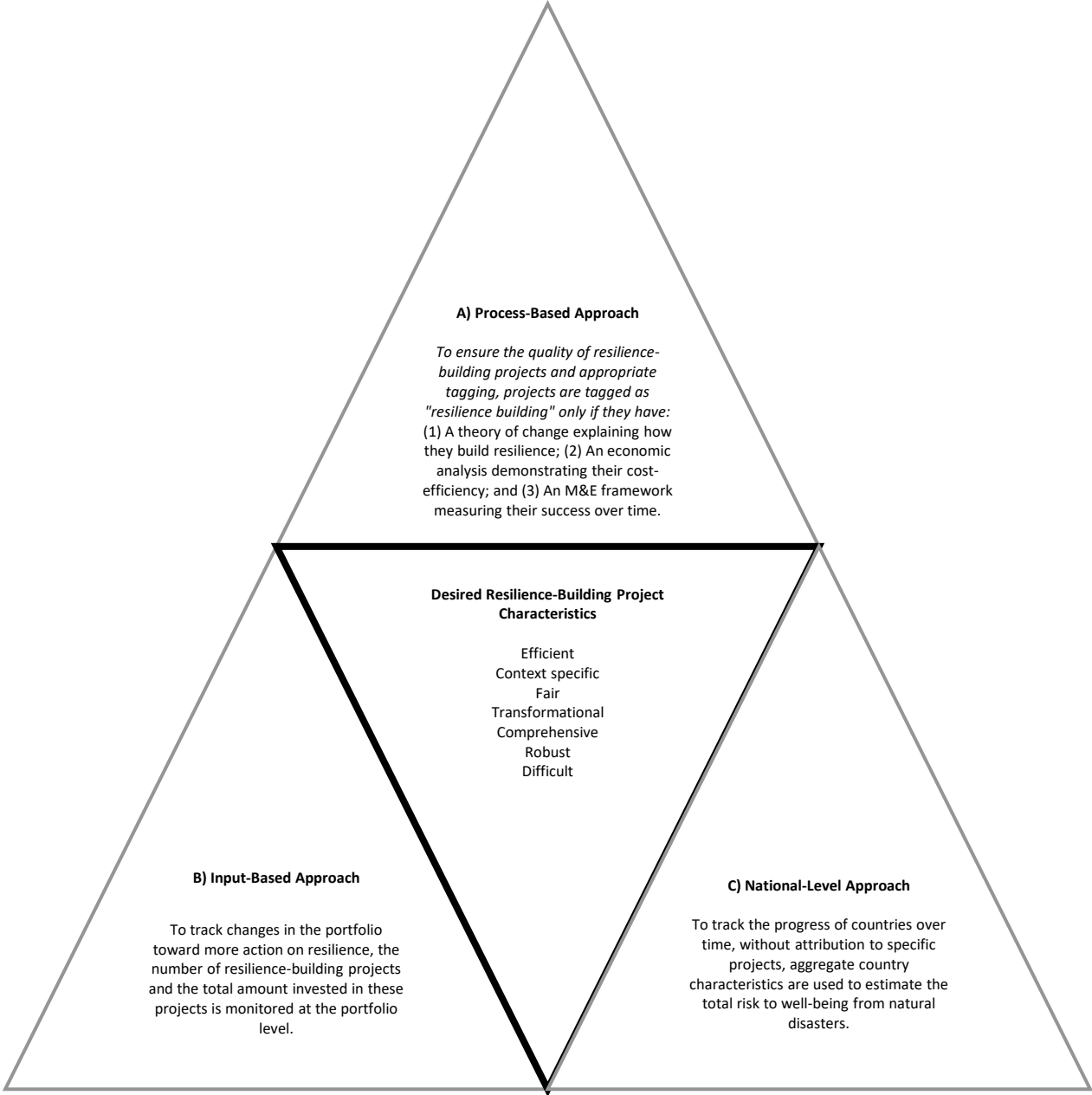
What can be the way forward? One option is to combine three complementary approaches, as illustrated in Figure 1: a *process-based approach* to ensure the quality of the portfolio and appropriate outcomes are produced, monitored, and reported; an *input-based approach* to track changes in the portfolio toward more action on resilience; and a *national-level indicator* to track the progress of countries over time.

- **Because resources provided for resilience building have to be used efficiently, one needs to be rigorous and strict on which projects are tagged as “resilience-building,” and to reinforce robust M&E of these projects.** In practice, every project that is tagged as resilience-building should have (1) A clear theory of change (how is the project supporting resilience? What is the “pathway” to resilience that is promoted?); (2) A solid economic analysis showing the project is achieving its objectives efficiently (at the lowest possible cost); and (3) An M&E framework that reports, over the lifetime of the project, on a set of indicators that measure the success of the

project in enhancing resilience. These indicators would be context- and project-specific, based on the chosen pathway to resilience, and should take into account as much as possible the designed characteristics of resilience-building projects (i.e. they need to be efficient, context specific, fair, transformational, comprehensive, robust, difficult). For instance, if the project is building resilience by driving new urbanization toward safer places, the fraction of the population living in the flood zone can be used as an indicator of success. The benefit from this approach is that it ensures an efficient use of resource used to boost resilience (thanks to the economic analysis) and the ability to assess ex post that the project had the impact that was envisioned (thanks to the M&E framework), while maintaining the flexibility needed to select the best projects (because indicators are project specific). The unavoidable drawback of this approach is that it does not allow aggregating the resilience benefits of multiple projects into a single number that would measure the resilience benefits generated by a portfolio of projects, since each project has its own set of indicators.

- **Because it is important to track changes at the portfolio level, which requires aggregation across projects, one has to use input-based indicators, such as the number of resilience-building projects and the total amount invested in these projects, since outcome-based indicators cannot be aggregated.** The drawback of an input-based metric would be largely mitigated by the requirements that projects need to meet to be tagged, as described earlier: (1) Using an aggregated input-based indicator (vs. an outcome-based one) is appropriate if there is insurance that each project has a strong M&E, and outcome indicators are measured and tracked at the project level. (2) Negative side-effects of input-based indicators (for instance favoring expensive solutions over cheaper ones) are mitigated by the strong economic analysis that demonstrates the cost-efficiency of the pathway to resilience that has been chosen and its appropriateness to the context of the project.
- **Because it is useful to track countries' progress over time, the projects and portfolio indicators can be complemented with a country-level indicator, such as the World Bank's socio-economic resilience indicator or the corresponding estimate of the risk to well-being (which includes the dimensions related to exposure, vulnerability, and socio-economic resilience).** Country-level indicators do not allow attribution of progress (or lack of progress) to a project or a set of projects, but it helps identify good practices (by investigating countries making rapid progress) and countries where more needs to be done to boost resilience.

Figure 1. Complementary metrics and indicators to measure the contribution of projects and programs to resilience



These resilience metrics should be considered within the right framework, in which resilience is a way of achieving prosperity, reducing poverty, and increasing the quality of life; not an end in itself. Even if resilience could be perfectly measured, not all projects that increase resilience would be desirable. For instance, driving farmers toward low-productivity and low-risk crops would increase their resilience, but

possibly at an unacceptable cost for their average incomes. Additionally, some resource-rich countries are highly resilient to natural disasters because a large fraction of their population survives off of transfers from the government, but they do not contribute to economic activity because they lack jobs. Since their livelihood cannot be damaged by a local event, these people may be very resilient, but at the expense of their ability to contribute to society and of the growth potential in the country. To avoid “overvaluing” resilience, it is important to consider the resilience benefits from a measure of projects in a broader framework in which other dimensions – such as average income and long-term prospects – are also included.

Finally, and most importantly, the limits of the indicators and monitoring systems need to be clearly communicated and accounted for in the decision-making process. Like with any (unconscious) bias, being aware of the bias created by quantified indicators is a critical first step to manage these risks and prevent imperfect indicators from leading to bad decisions. Hopefully, the thought experiments proposed in this note will help communicate the limits of potential resilience indicators that an institution might pursue, and will ensure that decision-makers take these limits into account so that resilience-building actions can deliver the maximum level of benefits, as much as possible.