This article discusses the dramatic changes that will face evaluation practice and the “culture of evaluation” in Africa as the 4th Industrial Revolution (4IR) evolves. The drivers of the 4IR will be the rapid evolution of six “disruptive technologies”, five of which present great opportunities for evaluation practice, but also significant challenges. The culture of evaluation will be transformed, and the key questions are how well the transformation will be managed and whether evaluation will remain a major force in the new big data ecosystem.
The culture of evaluation

It is helpful to begin by defining what we mean by the “culture of evaluation”. Mayne (2008) argues that an evaluative culture denotes an organizational culture that deliberately seeks out information on its performance in order to use that information to learn how to better manage and deliver its programs and services, and thereby improve its performance. An organization with a strong evaluative culture engages in self-reflection and self-examination as well as evidence-based learning, and encourages experimentation and change. Mayne also discusses ways to develop an evaluative culture, focusing on leadership, organizational support structures and a learning focus.

While this provides a helpful framework, it is important to recognize that evaluations are conducted for many different purposes, by different kinds of organizations and at different levels – ranging from small projects in a few villages, to a 15-year global evaluation program for assessing the effectiveness of the Sustainable Development Goals (SDGs). Evaluations are also conducted by various organizations operating with different mandates.

Evaluations are likewise conducted for varied purposes, with agencies giving different priorities to differing purposes. A widely used classification identifies three main types of evaluation: developmental evaluation (designing and testing innovative approaches for addressing problems, particularly for understanding programs operating in complex contexts); formative evaluation (helping managers and planners to improve the design and implementation of an ongoing intervention or to learn lessons that can improve future interventions); and summative evaluation (assessing the overall merit, worth and significance of a program). However, evaluations may have a range of other purposes.

Another important dimension concerns the evaluation design. At least ten main types of evaluation design have been identified, each involving a unique methodology with differing sets of assumptions about broader considerations such as the nature of evidence and whether there are objective “facts” that remain constant for different observers (epistemology and ontology). These designs are listed in Box 1. The list is divided into current evaluation designs and new, evolving designs adapted to the new reality of the Fourth Industrial Revolution (4IR). However, there is a continuum and some current evaluation designs are starting to address these new realities.

So, given the fundamental differences between many of these methodologies, as well as the diverse purposes of...
evaluation and the varied agencies that commission, implement and use evaluations, it is an open question whether there is a single, all-embracing “evaluation culture” or a number of different “evaluation cultures”. This is an important question for the present discussion as the 4IR, and the growing importance of new information technologies, will have different implications for different kinds of evaluations, conducted by different organizations.

Many evaluation approaches focus on continuity and are resistant to change

An important element of what could be called “mainstream evaluation” (the approaches used by many United Nations, multilateral and bilateral development agencies and which are taught in many evaluation training programs) are the demands for standardized evaluation approaches which are widely applied and which only change slowly over time. There are a number of factors influencing continuity.

Methodological considerations: Mainstream evaluation theory argues that the best way to assess causality and impact is to obtain a baseline measurement on outcome indicators and then to repeat the measurement using the same sample and indicators at the completion of the project. Ideally, the same data should be collected for a matched comparison group. This methodology requires that identical indicators be used at different points in time. Agencies that are required to evaluate all their programs will often invest significant resources in developing and testing the measurement instruments and in

<table>
<thead>
<tr>
<th>BOX 1. CATEGORIES OF EVALUATION DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current evaluation designs</strong> (all of which can be adapted to the 4IR)</td>
</tr>
<tr>
<td>1. Experimental designs</td>
</tr>
<tr>
<td>2. Quasi-experimental designs</td>
</tr>
<tr>
<td>3. Statistical designs (econometric)</td>
</tr>
<tr>
<td>4. Theory-based evaluations</td>
</tr>
<tr>
<td>5. Case-based evaluation – including Qualitative Comparative Analysis</td>
</tr>
<tr>
<td>6. Participatory and qualitative evaluation</td>
</tr>
<tr>
<td>7. Gender responsive evaluation</td>
</tr>
<tr>
<td>8. Synthesis and review</td>
</tr>
</tbody>
</table>

selecting the sample. Consequently, there are incentives to continue to use the same instruments to ensure comparability of results. To a critic this can be seen as a conservative mentality, which is resistant to introducing new instruments and indicators to adjust to change.

**Organizational and institutional considerations:** Many agencies conduct large numbers of evaluations of similar programs and need a standardized approach to reduce time and cost as well as to ensure quality. For example, agencies that work at the global or regional level are required to conduct periodic evaluations of their country programs, or sometimes to evaluate each of their individual loans/investments. Often evaluations will be sub-contracted to consultants, and consequently standardized methodologies, procurement procedures and terms of reference may be required.

For some national evaluation agencies, transparency is a priority. Particularly where evaluations are used as part of government decision-making to determine resource allocation among different states or communities, it is important to develop consensus on the methodology, and to provide stakeholders with access to the methodology.

As we will discuss below, these methodological and organizational factors contribute to the slowness of many evaluators to adopt potentially powerful new tools and techniques for data collection and analysis that are becoming available through the 4IR.

**The culture of evaluation in the 4IR: The need for a new evaluation paradigm**

A 2019 AfDB study, “Unlocking the potential of the 4th industrial revolution in Africa,” identified six potentially disruptive technologies that will dramatically change economic development patterns in Africa: Artificial intelligence (AI); the Internet of Things (IoT); Big Data; Blockchain; drones and 3D printing. In parallel to the dramatic transformations that these technologies are already producing in industry and commerce, they are also beginning to influence how development programs are designed, implemented and monitored.

All of these technologies dramatically increase the kinds of data that are available for the design, implementation and monitoring of social and economic development programs, and potential reduce the costs of access to and analysis of these data. This includes the more precise targeting of products and services to particular audiences, and the early detection of problems with project implementation of intended outcomes.

The rapid evolution of these information technologies and the increasing range of their applications has very important implications for development evaluations. There are three main challenges:

- How to assess the impacts of these new information technologies on economic and social development – including who benefits and who is excluded or may even be worse off;
- How to assess the impacts of these technologies on development policies and programs; and
- How to assess where, to what extent and how these new technologies can be incorporated into the design and implementation of development evaluations.

Initial impressions suggest that these kinds of questions concerning evaluation receive very little attention in many discussions on the 4IR. For example, the 2019 AfDB study mentioned above focuses on the business case for the promotion of these new technologies but does not discuss...
how to evaluate the effectiveness and impacts of these technologies, nor is there any discussion of unintended consequences. We will argue in the following sections that the implications of these new information technologies are so profound that they will require a paradigm shift in the culture of evaluation.

The challenges facing most current evaluation approaches, and which will be exacerbated by the 4IR

Evaluators are aware of the methodological, economic and organizational challenges that already confront the evaluation of current development programs. Many of these challenges will become more critical for evaluation in the 4IR era, and other new challenges will emerge. Some of the current evaluation challenges, all of which will continue in the 4IR era include:

Cost of data collection: For many evaluations, data collection represents the largest cost. Each interview involves a significant cost, therefore creating pressure to reduce the sample size or to exclude difficult to reach groups (such as people living in remote areas). The small sample size reduces the statistical precision of the analysis and often makes it impossible to compare different population groups (male and female-headed households, age groups, kinds of agricultural production, etc.).

Contextual analysis: Program outcomes are affected by economic, political, ecological and other factors in the areas surrounding the project. Similar projects may have very different outcomes depending on these contextual factors, but the cost of data collection usually limits a systematic assessment of these factors.

Addressing complexity: While evaluators and other stakeholders frequently state that their programs are complex, very few evaluations are able to address complexity in a systematic way. This is partly due to the large amounts of data required for a complexity analysis, but also because it will usually be necessary to incorporate systems analysis – which often requires the use of dynamic modeling and real-time data, which usually cannot be generated through conventional evaluation methods. The evaluation of the SDGs is an example of an evaluation that ideally requires a complexity focus and the use of systems analysis, but which are usually not possible.

Sustainability and longitudinal data: Most programs are designed to continue to operate, delivering services over a relatively long time horizon. However, for budget and organizational reasons, most evaluations only cover the period during which a program is being implemented and perhaps one or two years after it becomes operational. So, in most cases almost no information is obtained on whether the program and its benefits were sustainable. Furthermore, very few evaluations are able to collect information on the years before the project begins.

We will argue that the new information technologies that are becoming available under the 4IR offer exciting new ways to address these issues.

While all of the above challenges will continue, many 4IR development programs will face new evaluation challenges, including:

- Many programs are becoming larger and more complex.
- Programs will be more dynamic and rapidly changing.
- The private sector will be recognized as a major development partner and public-private partnerships will become more central to development.
- Increasingly large volumes of data will be generated more rapidly, from
multiple sources and of many different kinds (e.g. satellite and drone images, numerical, audio-visual, text); these will need to be captured and processed.

- Broad development challenges such as climate change, refugees and unregulated migration.
- The SDGs pursue 17 ambitious, complex and inter-related goals; as discussed above, these provide major challenges for evaluators.
- Increasing demands from governments, donors, civil society and other stakeholders for accountability and for greater access to information on public sector interventions.
- New information technologies are becoming accessible to any organization or individual with access to a computer and a certain level of technical expertise; consequently, governments and donors will no longer have a monopoly on information about their programs.
- Pressures towards social inclusion and popular participation.

Oppportunities of the 4IR: Big data and data science offer powerful new tools that can potentially strengthen development evaluation

As well as presenting challenges for evaluation practice, the 4IR also comes with opportunities. Big data is one such opportunity. Big data is often defined as data that is too large to handle on a normal office computer. It is generated very fast, produces huge volumes of multiple kinds of data, and often generates continuous data over long periods of time. Big data will be discussed in more detail in a second article, but the following are some of the exciting benefits for development evaluation in the 4IR era:

- Data is cheap and fast to collect.
- Very large volumes of data can be collected, often covering the total population of interest, not just a small sample.
- Data can be analyzed very rapidly and can provide continuous feedback.
- Evaluators can combine multiple sources of data (surveys, text—including huge PDF files, audio-visual data, satellite and drone images, photographs) into an integrated data platform.
- Evaluators can have much more sophisticated kinds of analysis including predictive modeling, machine learning and artificial intelligence, the integration of multiple data sets, and the analysis of complex programs.
- Analysts can generate continuous records, including for the period before a program begins and after it ends.

Box 2 gives examples of big data sources that could be used to strengthen the evaluation of development programs in the era of the 4IR. The examples go beyond the “disruptive technologies” identified in the AfDB 2019 paper. Potential applications of these techniques will be discussed in the second article.

Challenges for integrating the new big data tools into development evaluation

There are a number of factors likely to affect the speed and extent to which 4IR data science tools and techniques are adopted by evaluators.

Different research frameworks and tools. One important challenge is that evaluators and data scientists use varying frameworks and tools, and consequently many evaluators are not familiar
The culture of evaluation in the age of big data: The need for a new evaluation paradigm for the 4th Industrial Revolution?

Data scientists make extensive use of real-time data (e.g. posts from twitter and other social media, the intensity of electric light in a community as an indication of the level of economic activity), while evaluators are more familiar with data generated from surveys, project records and statistics. As a result, many evaluators are concerned about data scientists’ perceived lack of concern about issues relating to bias, data quality and construct validity.

There are differences with respect to the role of theory. While many evaluators consider theory-based evaluation (e.g. theory of change) to be critical in the design of an evaluation and the interpretation of the findings, many data scientists either question the need for theory in the analysis of big data or they have a different approach.

One area of debate concerns the difference between causality and correlation. While the use of comparison group designs is a central part of many kinds of evaluation methodology, data scientists often work with large correlation matrices, sometimes arguing that with a large enough number of observations and correlations it is not necessary to discuss attribution and causality.

Another area of debate concerns data scientists who claim that they...
have complete population coverage whereas evaluators normally work with relatively small samples. Evaluators reply that data science is rarely able to achieve complete population coverage and that the groups left out tend to have important characteristics (such as being poorer) so that these exclusions may result in serious bias. Evaluators also argue that data scientists are often not concerned with bias and do not have a systematic approach to estimate exclusions and their impact.

Weak institutional linkages between data centers and evaluation offices. Based on a limited number of interviews with large development agencies, it seems that the linkages between data centers and evaluation offices are often not well established and many data analysts do not see support for evaluation as one of their primary areas of activity. Similarly, many evaluation offices have not taken the initiative to develop relations with a data center.

Limitations on access to big data. Another practical challenge is that many organizations, particularly those that are smaller and less well funded, may only have limited access to many kinds of big data.

Limitations of integrated capacity development programs. Many evaluation training programs do not include data science methodology, and similarly many data scientists have never received training in evaluative thinking and methods.

Adapting development evaluation in a world of 4IR – rethinking the culture of evaluation.

The era of the 4IR offers powerful new tools for the collection (e.g. big data, satellites and drones, IoT and blockchains) and analysis (e.g. predictive analytics and artificial intelligence) of the increasingly complex development programs that are likely to evolve. These tools make it possible for evaluation to work with larger and more disaggregated data, to broaden the geographical coverage to include contextual analysis, to incorporate vulnerable and difficult to reach groups, and to use systems analysis and other tools to analyze programs operating in complex environments. It also becomes easier to study behavioral change, attitudes and group dynamics.

At the same time, interventions to be evaluated are likely to become larger, multidimensional, interconnected, dynamic and constantly changing. These interventions will also affect and be affected by global challenges such as climate change, growing refugee problems, complex migration patterns and human trafficking. The powerful resources mobilized by 4IR actors also mean that multiple unintended outcomes must also be evaluated, as well as the “dark side” of the 4IR such as fake news and new ways to manipulate consumers and voters and to create social divisions. The sources of potential biases built into remote decision-making algorithms must also be assessed.

For the reasons discussed above, the current evaluation culture and the most common evaluation methodologies are not well suited to fully understand and evaluate this new world. As a result, we are left with the question as to whether the different factors that make evaluation resistant to change can be overcome and allow a new evaluation culture to evolve. Two alternative future scenarios for development evaluation can be imagined. In the utopian scenario, evaluation stakeholders and evaluators will adapt to take advantage of the powerful new tools that are becoming available, and a new and vibrant evaluation culture will evolve that will become a powerful player in guiding the evolution of development programs. There is also a dystopian
(pessimistic) scenario in which evaluation culture is resistant to change, does not take advantage of the new resources and becomes increasingly marginalized and replaced by new 4IR management decision-making tools.

In conclusion, we argue that there is an urgent need to completely rethink current evaluation methodologies to take full advantage of the new information technologies available. This allows evaluators to prepare evaluation agencies and address the challenges of assessing the more complex development strategies and programs that are already evolving. The opportunity is there to adapt the current evaluation culture(s) to the demands of the 4IR world. Will the evaluation community seize the opportunity?

Coming in Evaluation Matters Q2 2020: Part 2 - Transforming evaluation in the 4th Industrial Revolution: Exciting opportunities and new challenges

Part 2 of this article, coauthored by Michael Bamberger and Pete York, will appear in the 2nd Quarter 2020 edition of Evaluation Matters. The article will discuss how evaluation practice will be affected by the five “disruptive elements” of the 4IR, the exciting opportunities that these offer for evaluation, as well as the methodological, organizational and political challenges they will bring. It will also review cutting-edge evaluation technologies, which are already evolving in industrialized nations that will become available throughout Africa in this new era; and will look at the disruptive challenges that evaluators will be faced with during the process of transformation. One or more case studies will illustrate how the new evaluation approaches are being, or could be, applied to the different kinds of evaluations that the African Development Bank conducts.
Endnotes

1. A number of authors have also identified the following reasons for conducting an evaluation: (1) assessing the broader development impacts of a program (for example, contribution to one or more of the SDGs); (2) adapting an intervention to a new context or to a change in the policy environment; (3) contributing to resource allocation decisions on alternative programs; (4) helping to identify emerging problems and building consensus on the causes of a problem and how to respond; (5) supporting public sector reform and innovation; and (6) contributing to global development initiatives (MDGs) and challenges (climate change, human trafficking). Source: Bamberger and Mabry 2020 combining ideas from various authors.

2. An industrial revolution is a fundamental technological transformation that transforms the rest of the social and economic system and where there is a strong interdependence of the different elements of the system. Economic historians have identified three industrial revolutions: The First industrial Revolution (1760-1840); The Technological revolution (Second Industrial Revolution: 1840-1920); and The Digital Revolution (The Third Industrial Revolution: 1975-2020). The fourth industrial revolution, which does not yet have a universally accepted name, is based on dramatic changes in information technology and how it is applied and involves the blurring of boundaries between the physical, digital and biological worlds.

3. The complexity literature identifies at least four complexity dimensions that affect program design, implementation and outcomes: (1) the complexity of the program itself, (2) complex interactions among multiple stakeholders, (3) the systems within which the program operates and (4) the nature of non-linear causality (Bamberger, Vaessen and Raimondo 2016 Dealing with complexity in development evaluation).

4. in Evaluation Matters Q2 2020 edition, forthcoming

References


Michael Bamberger has a Ph.D. in Sociology from the London School of Economics. He has been involved for over forty years in the evaluation of development programs in Africa, Asia and Latin America, including 25 years as a senior sociologist with the World Bank. His work has focused on poverty and social exclusion, gender equality, urban development and the challenges of evaluating the MDGs. Over the past few years he has worked on the opportunities and challenges for integrating new information technology into the evaluation of development programs. His recent publications include: “Evaluation in the age of big data”; [with Jos Vaessen and Estelle Raimondo] “Dealing with complexity in development evaluation”, [with Linda Mabry] “RealWorld Evaluation: working under budget, time, data and political constraints”, and “Evaluating the Sustainable Development Goals through equity-focused and gender-responsive evaluations.” Over the past 20 years he has consulted with 10 UN agencies, multilateral and bilateral development agencies, development banks, foundations, NGOs, and governments on the evaluation of development policies and programs.