THE EES NEWSLETTER

EVALUATION CONNECTIONS

Editors: **Ole Winckler Andersen and Kevin Williams.** Guest editor: **Christiane Kerlen.** Editorial assistant: **Eva Petrová.**

EES EUROPEAN EVALUATION SOCIETY

MAY 2018

ARTICLES – NEWS – EVENTS

CONTENT

EDITORIAL: PERSPECTIVES ON THE FUTURE OF EVALUATION

Christiane Kerlen and Ole Winckler Andersen

Over the last few decades evaluation worldwide has undergone enormous changes, a fact reflected in the growing number of national and regional evaluation networks. One of the older evaluation societies is DeGEval – the Austrian and German Evaluation Society – which celebrated its 20th anniversary in 2017. Events like these often provoke a tendency to look back, but the Board of DeGEval chose a forward looking topic – 'The Future of Evaluation – Evaluation for the Future' – for its annual conference in September 2017. The aim was to stimulate a discussion of the threats, challenges and opportunities evaluation is facing in times of increasing demand as well as a heightened focus on evaluation results. The discussions at the conference have been summarized in a Position Paper by the DeGEval Board[1].

DeGeval invited EES to be part of the conference and to organize jointly a session that would introduce and discuss international perspectives on the future of evaluation. The first two articles in this issue of *Connections* are written by participants of this session.

Jan Hense, president of DeGeval, points to concerns regarding the future demand for evaluation. He argues that with the rise of social media, a common point of reference necessary for an informed public discourse is rapidly falling apart. The negation of expert opinions and the belief in fake news are a second trend contributing to a potential pessimistic view on the demand for evaluation. In his conclusion he emphasizes the important role of external forces in determining the future of evaluation.

Reinhard Stockmann and Wolfgang Meyer remind us of different indicators showing the success story of evaluation. Their outlook for the future takes its starting point in observing that evaluation demand is to a great extent policy-driven. They map out those societal context factors that predominantly influence the willingness of political elites

1 DeGEval Board (2017): The Future of Evaluation – Position paper 10 of the DeGEval. Position Paper of the Board of DeGEval – Evaluation Society. November 2017. www.degeval.org/en/publikationen/positionspapiere.

to use evaluations. One risk they see is that a push in standards by international aid organizations may result in a trend towards the homogenization of evaluation that is somehow alienated from national political cultures and the majority of societies.

In her article Karen Jorgensen discusses how the sustainable development goals (SDGs) and new models of providing development finance are influencing evaluation. She points to several challenges that result from multiple goals, fragmentation and rising complexity on a range of dimensions; for example, due to an increased private sector involvement in development co-operation. She concludes that the time has come to assess whether the criteria and standards that have guided development evaluations for a number of years are robust enough to respond to the new challenges.

In the fourth article Per Øyvind Bastøe and Siv J. Lillestøl discuss the link between evaluation institutionalization and use. The article mentions the increased institutionalization of many evaluation functions, but argues that a similar increase in the use of evaluation has not taken place. The authors point to a rhetoric-reality gap – also mentioned in Hense's paper – as one explanation. Another explanation is the complexity of many development programmes. One of the paper's closing reflections is that use is often too narrowly understood and that a broader understanding of use would be relevant.

The final article by Ole Winckler Andersen compares five recent meta-evaluations and identifies both methodological similarities and differences. Examples of differences are size of samples and number of indicators. The article concludes by identifying several issues which could be considered and discussed in the context of future meta-evaluations, including the use of criteria and rating procedures.

■

GERMAN ANGST REGARDING THE FUTURE OF EVALUATION

Jan Hense

In some regards, making predictions about the future is one of our core tasks as evaluators, though we may not always be aware of it in our daily routines. In doing evaluations, just as in any other profession working with empirical data, we are inherently forced to draw on past events: each evaluative conclusion originates as a conclusion about the state of things at the point of data collection. However, most evaluation consumers would rightly be frustrated if our reporting was limited to the past. They will most certainly expect certain predictions about the future. All recommendations are a gamble that certain consequences will occur in the future if they are implemented. Similarly, our evaluative conclusions about an evaluand, although derived from past events, often imply the prediction that the same evaluand in the future will bring forth similar results, particularly in regard to its outcomes. A third area where evaluators aim to predict the future is ex-ante, preformative, or input evaluation, the latter in the sense of Stufflebeam's (2000) CIPP model. Even if these types of evaluation draw only on past data, our task as evaluators implies a prediction of the extent to which it will be worthwhile to implement a given plan, concept, or proposal in the future.

Given these examples, it would seem that evaluators are well equipped with the prophetic powers necessary to answer questions about the future fate of their own discipline. Alas, this is not the case. Contrary to predictions in most evaluation contexts, here the scope of the question is too wide, the terminology too ambiguous, the number of influences too large, their interrelations too complex, and previous experience too scarce, for one single person to be able to make more than an educated guess. However, what I can do is to share some current concerns about the future demand for evaluation.

There are a number of reasons on different levels contributing to the demand for evaluation. Curiosity may be a factor at the individual level, conformity with regulatory regimes a reason at the institutional level. However, in the context of globalisation and society, at least two main drivers are at work. One is the need to provide justification for public policies in democratic societies, the other is the conviction that facts and evidence can and should play a crucial role in decision making. It seems that both of these drivers have recently come under serious attack as a more or less direct result of factors commonly attributed to globalisation.

The need to provide justification for public policies depends on a strong civil society and critical discourse in the media. Given the ongoing fragmentation of the media landscape as a byproduct of the rise of the internet and social media, accompanied by the emergence of echo chambers and filter bubbles, it seems that the common point of reference necessary for an informed public discourse is rapidly falling apart. As a consequence, there is reason to fear that the demand to justify public policy with the help of evaluation will wane, not because citizens will stop asking for justification, but because there is not enough common ground among the citizenship concerning the construction of what constitutes social reality. Why call for evaluations if there is no longer a coherent narrative about a social problem that their findings can contribute to? Of course, it might be argued that such an idealised state of common social co-construction of reality has never really existed and that societies have always been fragmented in regard to individuals' perceptions of the world. However, it seems that the necessary overlaps between these perceptions that are needed to provide a critical mass of common understanding are rapidly being lost.

This leads to our second driver, the belief in the role of evidence in decision making. Here, also, one should not be so naïve as to think this role has been easy in the past.

Hard-learned lessons taught evaluators that evidence is just one factor among many when it comes to individual and collective decision making processes, and that knowledge generation is always embedded in and influenced by historical and social contexts. Accordingly, as Cronbach observed quite early in the profession's development, "the notion of the evaluator as a superman who will make all social choices easy and all programs efficient, turning public management into a technology, is a pipe dream." (Cronbach et al., 1980, p. 4). But even if the nature of evaluative evidence has never been undisputed, even if particular evaluative findings have always been confronted with questions regarding their validity, there used to exist a general belief that the scientific method of knowledge generation, as through systematic evaluations, is inherently apt to provide insights that should be influential in decision making. Yet, having become allergic to the complexities and ambiguities of living in a globalised world, it seems that more and more people no longer share this belief. Many are increasingly ready to refute scientific evidence as elitist, and replace it with fake evidence and recursive self-confirmation of preexisting world views. Here also, we are witnessing a landslide shift regarding a phenomenon that may have always existed. Today, the belief in conspiracy theories and fake news is no more exclusive to the casual oddball. It has arrived at the highest levels of decision making of western democracies as an influence and a tool of policy making at the same time. What role can evaluation play in an arena where its foundational idea of using scientifically driven, systematic, and transparent methods to arrive at insights about the merit or worth of a social practice is not accepted anymore?

In spite of these pessimistic perspectives, many positive developments within the field of evaluation can also be found in recent years, and many of them due to globalisation as a driving influence: transnational interchange and dialogue; effortless means of worldwide communication and cooperation; easy access to exhaustive data sources; new means of data analysis; numerous theoretical and methodological advances; more and better research on evaluation; and a general trend towards more professionalisation. So, as a field a lot is being done to keep evaluation practice moving forward. But there is an elephant in the room. The fate of evaluation will be determined by external forces. It relies on the need to justify public policy within civil society and belief in scientific methods. If too much ground is lost to the enemies of an open society, it will no longer matter how well evaluators do their work.

References

Cronbach, L. J. et al. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.

Stufflebeam, D. L. (2000). The CIPP model for evaluation. In *Evaluation models* (pp. 279–317). Springer Netherlands.

EVALUATION BETWEEN EVIDENCE-BASED POLICY AND "FAKE NEWS": PATHS TO THE FUTURE

Reinhard Stockmann and Wolfgang Meyer

On one hand, prospects for the future of evaluation are very promising. Worldwide, state and non-state demand is booming. The range of evaluation services offered by consulting and scientific institutions, as well as teaching and training opportunities, continue to grow almost everywhere. On the other hand, there are also dangers lurking for the evaluation boom. The global rise of populism is accompanied by self-produced "fake news", which represents a clear attack on evidence-driven policy and its tools like monitoring and evaluation.

The global rise of evaluation culture and practice

There are several indicators to demonstrate the success story of evaluation, especially in North America, Europe, Australia and South Africa (Meyer & Stockmann 2016; Meyer 2018):

(1.) In many countries, evaluation is a *fixed element in policy-shaping* and a management control element in international organisations, national governments and their administrations, and a wide range of non-profit organisations (Rosenstein 2015). Datta (2006: 420) points out that 'scientific-research-based programs and evaluations … have become widely institutionalized for all manner of programs.'

(2.) The number of *national evaluation societies* has grown considerably in recent years. According to a worldwide Internet search by Dahler-Larsen (2006: 142), the number of evaluation societies increased tenfold to 83 between 1984 and 2004. The strongest growth in recent years has been in Europe and Africa (Meyer 2018). Currently a total of 150 organisations are listed on the website of IOCE – "International Organization for Cooperation in Evaluation", including 122 at national and 12 at international level.[1]

(3.) Increasing demand has given rise to a *broad demand market* for evaluation, which is continuing to grow (Leeuw, Toulemonde & Brouwers 1999: 487ff.). The number of consulting firms involved in evaluation has risen sharply. Apart from consulting companies, there are a number of research institutions and universities active in the evaluation mar-

1 www.ioce.net

ket and attempting to combine research and evaluation for their clientele, as well as basic and advanced training and communication in a fruitful way.

(4.) Above all, thanks to the development of information and communication technologies and the World Wide Web, the *dissemination of evaluation findings* has seen a tremendous surge. Even if many organisations still do not make their evaluation reports accessible to the general public, a host of findings from evaluations are now already available on the Internet.

(5.) *Training activities offered worldwide* have increased sharply. Both in North America and Western Europe, a broad variety of training and advanced training programmes and university study programmes is offered (Friedrich 2016, Meyer 2016). However, these training activities are not stable over time and only a few regular study programmes with more than 60 ECTS on evaluation have been permanently established.

Furubo, Rist and Sandahl (2002: 7ff.) made a first attempt to characterise evaluation culture in a comparative way. They applied nine criteria and a marking system to generate a ranking list. At the beginning of the 21st century, only the USA achieved the maximum possible number of points. Canada, Australia, Sweden, the Netherlands and the United Kingdom followed. Since then, professionalisation has progressed to very different degrees in the selected countries. In a replication of the study ten years later, Jacob, Speer & Furubo (2015) found a dramatic increase, especially in those countries that lagged behind the others (Japan, Israel, New Zealand, Spain and Italy), while the leading countries (USA, Canada, Australia and Sweden) slightly lost ground. Stagnation characterises the situation in leading European countries (United Kingdom, Netherlands and Germany), with the gap between the pioneers and the others closing during the last decade.

However, this research does not cover the most interesting development in evaluation culture because it is strictly limited to North American and Western European countries (with only a few exceptions e.g. Australia, Israel, Japan, New Zealand and South Africa). There has been a dramatic rise of evaluation practice in Africa, Asia, Central Eastern Europe and Latin America, guided by "good governance" concepts and evaluation cultures in multilateral organisations such as the UN, OECD, EU and the regional development banks. Countries like Costa Rica, Malaysia, Poland, and Uganda have become important hubs and driving forces for the diffusion of evaluation into the global regions that had been left behind in the 20th century and the early development period of evaluation. Moreover, there is some evidence for giving a leading role to countries from the global south in discussing evaluation as a tool for guiding the national implementation of the globally agreed sustainable development goals (SDGs) (Simon et al 2017, Meyer et al 2018).

To sum up, global equalisation characterises evaluation development nowadays. Differences between countries are becoming smaller and less relevant as the pace of evaluation development in the leading countries has slowed down, while in other countries evaluation "take-off" has been accelerating. But this does not necessarily lead to coherence and a single "global culture" of evaluation as a careful look into the future may reveal.

Future challenges and pitfalls

If one wants to take a look into the future of evaluation (Stockmann 2015), it is necessary to explore the societal framework conditions that made the developments of the past possible – and to investigate changes and trends. A look at historical developments demonstrates that evaluation is clearly policy driven (cf. Stockmann 2008 and 2013). Both the first boom in evaluation the 1960s and 70s in the USA and somewhat later in most countries in Europe, and the second boom which began in the 1990s, as well as the new developments in the Global South, were triggered by increased state demand. In other words the question of whether evaluations have to take place or not, whether or not the market for evaluation grows, stagnates or shrinks, and even what subjects are covered, are, to a great extent, politically influenced. That is to say, influenced by the willingness of clients to deploy funds for evaluation. Even if funding comes from international organisations and the global community, political willingness to improve policy performance and to use evaluation as a tool for rational decision making is a necessity at national level.

Some societal context factors influence the willingness of political elites to use evaluations today:

(1.) Firstly, the *increasing importance of global issues* (e.g. climate change, the financial crisis, migration) go beyond national frontiers and require joint actions to solve problems and a transnational network of control mechanisms. Because of these developments, the boundaries imposed by national evaluation cultures need to be overcome, the path for transnational joint evaluations to take place cleared and the functions of evaluation redefined. The alternative pathway is to deny the existence of shared global challenges and to relaunch competitive national egoism. This is exactly what President Trump and some other populist leaders in the World are doing, and their way of thinking is becoming increasingly successful.

(2.) Secondly, *concepts like evidence-based policy and New Public Management* strengthened the requirements for a rational approach to public policy and its impacts. The increasing demand for evaluation was a result of this. However, political elites have not followed this path and consequently success has been rather limited. More and more people are taking a populist approach that does not bother about facts and produces its own facts as "fake news" in a non-rational way. Additionally, the populist movement criticises traditional sources of knowledge production as interest-driven and challenges their claim of being honorable, trustful and neutral. As a result, evaluation may become purely a means of consultation, losing its scientific background and roots, with sweetheart reviews, glorified opinion polls and satisfaction surveys likely to predominate as consultancies compete for contracts.

(3.) Thirdly, the necessity of a *more sustainability-oriented policy-making* that integrates economic, ecological and social issues requires a strengthening of the reflexive function of evaluation. The sustainable development goals (SDGs) give this way of integrative systems-thinking a more prominent role than before. Hence, this challenges the traditional way that responsibilities are distributed across different policy departments, which have their own culture and style. While this kind of departmentalized thinking has led to different ways of doing and using evaluations, integration of different evaluation cultures in various policy fields becomes increasingly a challenge for the evaluation community. Moreover, the expansion of evaluation into further policy fields and areas of activities may lead to an inflationary use of evaluations, providing reasons for reactance of stakeholders.

(4.) Fourthly, the *dissolution of political boundaries* (Beck 1992: 183ff.) describes a development which is characterised by an intensified interdependence of government and civil society actors and leading to expanding opportunities and forms of participation in political decisions and their implementation. While 'network society' (Castells 2010) may still be too strong a term and governance does not yet replace government (Bellamy & Palumbo 2010), there is a strong voice for more inclusion of civil society and for participatory decision making. For evaluation, this development is likely to mean an increased demand for network evaluations and participatory approaches, and a strengthening of the democratic and reflexive function of evaluation. Such a development accompanied by increasing demands for specific skills may lead to a situation where the training of evaluators cannot keep up with these demands and the requirements of customers. If evaluations are not carried out in a professional and appropriate manner, sponsors could turn away from the instrument of evaluation and look for other methods.

Conclusion

These changes will affect the interests of evaluation sponsors in national political and administrative systems. Without a doubt, these systems differ significantly between states and their capabilities to address some of these challenges. There might be a tendency for differentiation according, for instance, to the strength of national populist movements in politics, leading either to a decline or a more isolated version of evaluation. Conversely, dependency on transnational support and aid may pressure political elites to implementation evaluation in a way that conforms to international standards and reflects the evaluation culture of multinational organisations. This may lead toward a homogenisation of evaluation that is somehow alienated from national political cultures and the majority of society. In the way that the EU is blamed for being a bureaucratic block on national development, evaluation may become the target of polemicists and become discredited in some countries due to its lack of integration into the local culture.

Undoubtedly, there are certain risks for evaluation, whether it follows a more isolated national or sectoral route or a more inclusive and integrative pathway. In general, the evaluation community has developed to be very open-minded and dialogue oriented. There are strong forces in support of different ways of thinking and a fair discourse on different approaches to how evaluations can be done. This may help to face future challenges.

References

Beck, U. (1992). *Risk society: Towards a new modernity* (Vol. 17). Sage.

Bellamy, R. & Palumbo, A. (eds. 2010). *From Government to Governance*. Albingdon: Routledge.

Castells, M. (2011). *The rise of the network society* (Vol. 12). John Wiley & Sons.

Datta, L. E. (2006). The practice of evaluation: Challenges and new directions. *The Sage handbook of evaluation*, 419–438.

Friedrich, V. (2016). European University-Based Study Programs in Evaluation: Characteristics and Future Challenges. In *"The Future of Evaluation"* (pp. 113–128). Palgrave Macmillan, London.

Furubo, J. E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation*. Transaction Publishers.

Jacob, S., Speer, S., & Furubo, J. E. (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, *21*(1), 6–31.

Leeuw, F. L., Toulemonde, J., & Brouwers, A. (1999). Evaluation activities in Europe: a quick scan of the market in 1998. *Evaluation*, *5*(4), 487–496.

Meyer, W., & Stockmann, R. (2016). Conclusion: Shared Perspectives for a United World of Evaluation? In *"The Future of Evaluation"* (pp. 328–357). Palgrave Macmillan, London.

Meyer, W. (2016). Toward Professionalization? The Contribution of University-Based Training Programs in Pioneer Countries. In *"The Future of Evaluation"* (pp. 98–112). Palgrave Macmillan, London.

Meyer, W. (2018). Evaluation as Profession 2016. A global perspective. In *EES Connections*, January 2018.

Meyer, W., Naidoo, I., D'Errico, S., Hofer, S., Bajwa, M., Perez, L.A.T., El-Saddik, K., Lucks, D., Simon, B. & Piergallini, I. (2018). VNR reporting needs evaluation: A call for global guidance and national action. In *IIED-Briefing*, No. 8, January 2018.

Rosenstein, B. (2015). *Status of National Evaluation Policies Global Mapping Report*.

Simon, B., Meyer, W., D'Errico, S., Schwandt, T., Lucks, D., Zhaoying, C., El-Saddik, K., Schneider, E., Taube, L., Anderson, S.; Ofir, Z. (2017). Evaluation: a missed opportunity in the SDGs' first set of Voluntary National Reviews. *IIED-Briefing May*. Retrieved from http://pubs.iied.org/pdfs/17423IIED.pdf

Stockmann, R. (2008). Evaluation and Quality Development: Principles of Impact-Based Quality Management. Frankfurt et. al.: Peter Lang.

Stockmann, R. (2008). *Evaluation and Quality Development: Principles of Impact-Based Quality Management*. Peter Lang.

Stockmann, R. (2017). The Future of Evaluation in Society. In: Stockmann, R. & Meyer, W. (eds.): *The Future of Evaluation. Global Trends – New Challenges – Shared Perspectives*, Basingstoke: Palgrave Macmillan. ■

EVALUATION AND THE SDGS

Karen Jorgensen

The sustainable development goals (SDGs) and the new models of providing development finance are influencing how development cooperation is being carried out today and in the future. These changes also have an impact on development evaluations.

First, the sustainable development goals have implications for the way development cooperation must be carried out.

Secondly, a stronger emphasis on working with the private sector and in situations of conflict and fragility has clear links to foreign policy and national self-interest.

Thirdly, changing policies and priorities mean changes to patterns of co-operation – perhaps even signalling a farewell to long-term partnerships.

The new development agenda

Because the new development agenda is broad and the goals inter-related more institutions are involved in delivering the SDGs across government: evident when it comes to the national response, but also true for development co-operation. Delivering on the SDGs will take a whole-of-government if not a whole-of-society effort, affecting both how activities are designed and how evaluations take place.

Official development assistance (ODA) is increasingly spent by departments other than the ministry primarily responsible for development cooperation and ODA, typically the ministry of foreign affairs. Such "fragmentation" is driven in part by many developing – or partner – countries no longer needing broad-based, long-term technical assistance, but rather more pointed ministry-to-ministry cooperation to build specific capacity.

Challenges for evaluation

This dispersion makes it more difficult to plan development interventions, to have a concerted view of what is in the portfolio, especially at partner country level, to have common result and monitoring frameworks and to ensure evaluability. And it makes it more difficult to ensure evaluation across all ODA spending: who is responsible for evaluating the outcomes of interventions by other departments? If the spending is counted as ODA, should it be evaluated by the office charged with evaluating the core ODA budget? This is an issue many members of the OECD Development Assistance Committee (DAC) are dealing with, and it comes with the risk that substantial parts of the budget are not subject to evaluation.

A further complication is that not only are interventions more likely to cut across several policy fields to address the sustainable development goals, but interventions must also be sustainable on all three dimensions: economic, social and environmental. Currently DAC evaluation criteria define sustainability as being *concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn.* The criteria do mention that the projects need to be environmentally as well as financially sustainable. But the interconnected nature of the goals may make it more difficult to measure and evaluate and to add the social dimension.

Partnerships and evaluation

The SDGs also call for new partnerships and the Addis Ababa Action Agenda puts the role of the private sector squarely at the heart of how to achieve the SDGs. Many governments have – or are setting up – finance institutions to encourage private sector involvement in development co-operation. If these relate directly to a ministry or agency, it is relatively straightforward to set up monitoring and evaluation systems. But experience shows that the evaluation culture of such finance institutions, to the extend they have one, is not geared to looking at development outcomes. Bringing in the private sector does not make the equation any simpler as private sector companies will have different measures of success, and they are often not prepared to evaluate their investments citing protection of commercial interests. But if the enterprise has benefitted from ODA as credit-guarantee, risk protection, or through blended finance, is it not reasonable to ask that the interventions and outcomes be evaluated for their development impact? This is a question – and a challenge – development cooperation administrators need to address as more and more private sector instruments are rolled out.

In addition, the private sector may think differently about what sustainability means. A World Bank look at Public-Private-Partnerships showed, for example, that PPPs hardly ever addressed impact on government expenditure, i.e. the fiscal sustainability of projects. There is also already talk of "SDG washing": private sector companies marketing their products as contributing to attaining one of the goals while ignoring the impact they may have on other goals.

Multiple goals and evaluation

The deepened integration of development co-operation policy and foreign policy can also complicate life for evaluators, yet it is increasingly necessary to grapple with these issues, especially as more – and an increasing proportion of – ODA is spent in fragile contexts. In conflict and fragile situations, there are often mixed objectives for interventions. Of course, humanitarian concerns will play a key role for assistance in humanitarian emergency situations. But these can coincide with foreign policy and security interests which do not always pull in the same direction, at the same time, in the same space.

The difficulty for evaluators will be to disaggregate the objectives and the results. If a foreign policy objective has been met, but the development objective has not, can we declare success? In conflict and fragile situations it is difficult to get right the balance of power and influence in a whole-of-government approach: it is difficult to balance security interests with development interests. It is even more complicated when considering also the broader policy objective that development co-operation must meet national interest of the donor government. National interest can be hard to define and therefore it is difficult to measure –and evaluate – national interest outcomes along-side development ones. As most DAC members now have integrated foreign policy and development co-operation administrations, evaluation departments should increasingly be called upon to evaluate across the ministry.

Targeted interventions and evaluations

As more partner countries make headway in their development, they require less broad-based development support. And as donors' policies are changing, they tend to look to shorter-term, more specific interventions in partner countries.

This is leading to more targeted interventions, including more specific technical co-operation (which can be very useful). Keeping up with evaluating many smaller interventions, especially from a sustainability point of view, will be resource intensive and even if done, it may be hard to capture lessons which are meaningful without a broader development context. What can help here is stronger national institutions at the partner country level, another challenge which development cooperation evaluators have tried to address for decades and where a new approach is needed if there is to be real change and impact.

Reassessment of the evaluation criteria

It is now timely to assess whether the criteria and standards that have guided development evaluations for more than a decade are robust enough to respond to the challenges. The criteria of relevance, effectiveness, efficiency, impact and sustainability still seem appropriate, but do they need to be interpreted differently? The criteria and standards should not be regarded as a straight-jacket, but as a framework. For example, a reinterpretation of sustainability would be timely. Relevance may also be more complicated to determine: whose needs are considered most important when there might be competing interest meeting national interest with the needs of the partner countries; are development interventions addressing the needs of partner governments or are they aiming to meet needs of whole-of-society? The criteria and the standards are still relevant to guide evaluators in the new development context but having a debate about what that means would be useful. To this end the review started by the OECD DAC Development Evaluators' Network will be a good contribution.

Evaluation architecture and new technology

The increased attention to accountability for ODA spending has led to the setting up of independent evaluation institutions. Yet, there will still be a need for in-house evaluation capacity – capacity being a challenge in itself – and the question of independence is still critical. How can the evaluation function's independence be ensured while at the same time placing it close enough to policy makers to influence decisions through evidence?

The use of new technology is already changing how development cooperation impact is tracked, monitored and followed up. Change can be captured in visible data that are available continuously. There is more demand for real-time information, faster feedback loops, more adaptation. The emergence of new data sources (from communication devices, satellites, sensors) allows for more analysis.

Many evaluations currently take months if not years to complete. They often use surveys that offer a time-bound snap shot. But big data (for example, data generated through social media) show more patterns over time. It is possible to watch change in real time though the data will not reveal why change is taking place.

Evaluators will still need to be able to make complex processes of change easy to grasp and understand and new technology can add more information, but it will not fundamentally change the essence of evaluation. However, using new technology could lead to shorter feedback loops addressing the growing demand for actionable information earlier. There are some issues evaluators must consider as they begin to make use of new technologies and sources of data:

- Data availability: most big data are held by private companies and can be considered both highly valuable and potentially sensitive.
- Built-in bias of big data: not all relevant stakeholders will have access to the technology that generate big data.
- Capacity to use data and data analysis techniques: higher demand for evaluators who can meet tell stories about why change is happening. Such analysis requires more contextual understanding and not just number-crunching.

Expanding access to the internet in developing countries and greater connection to more diverse sources of information are leading to growing expectations that information will be publicly available in turn leading to growing demands on governments to be accountable.

Speaking truth to power

To contribute to accountability, evaluators must be willing to challenge assumptions, to take risks and be willing to look at contentious areas of policy, and to dare to examine in depth sensitive political issues and areas where development co-operation and governments are failing to deliver value for money. This requires a culture of evaluation across government. Evaluators can only contribute to accountability to the extent that they are able and empowered to look at difficult issues and speak truth to power, particularly when that truth is inconvenient.

For evaluators, the struggle between accountability, learning, and direction continues. Development cooperation programmes are evaluated in many cases because decision-makers – politicians – know they are accountable to parliaments and the public for the outcomes of development efforts. But evaluations are still in many corners seen as punitive – the fear of the negative – and not for the benefit they can bring.

What can evaluators do about the reluctance to evaluations coupled with shorter time-horizons and attention spans? Evaluations need to be timelier so that they are ready to offer up the evidence needed to influence policy or design of the next programme.

This will mean employing different – perhaps invent new – ways of evaluating, accepting that there are different thresholds of rigour for different purposes. Communicating in smarter ways using the new tools technology can offer – will continue to be on the agenda of evaluators.

But the challenges and opportunities set out above belong not just to evaluators: addressing them and taking full advantage will require that policy makers, programme developers and evaluators all contribute to setting the scene for better, more timely evaluations in a complex development landscape. ■

TOWARDS A GROWING GAP BETWEEN EVALUATION INSTITUTIONALISATION AND USE?

Per Øyvind Bastøe and Siv J. Lillestøl

Institutionalisation of the evaluation function

Recent studies of evaluation systems in UN organisations (UN, 2014) and in development agencies, ministries and multilateral banks (OECD, 2016) display a seemingly optimistic trend when it comes to institutionalisation of the evaluation function in development cooperation organisations.

One of the conclusions of the UN study is that the central evaluation function of UN organisations has changed with regard to roles, structure, systems and standards. It has moved from a predominant role of oversight and quality assurance of decentralised evaluations to focusing on supporting broad and strategic corporate-level decision-making. The OECD study, which looks at 45 evaluation entities, finds a related trend. Compared to a study in 2010, these organisations have now improved capacity and are more focused on strategic questions.

These findings resonate with a similar study of institutionalisation of evaluation in national governments (Jacob et al, 2015).

From our own perspective within the Norwegian government's development policy organisation, the same trends can be observed. In the government's financial management regulations, it is now mandatory to do evaluations of financial transfers and grants. The establishment of the Norwegian evaluation association in 2010 is another indication of a greater attention to evaluation. In the development policy administration, the mandate of the evaluation function was revised in 2015 to give it a more independent and prominent role.

Increased institutionalisation of the evaluation function is in line with a widespread public governance conviction that this function is important for transparency, accountability, efficiency and effectiveness of governments. This implies a belief that increased institutionalisation would translate into increased use – in different ways – of evaluation findings and recommendations.

Use of evaluation findings and recommendations

Use is commonly understood as the main purpose of evaluation (OECD, 1991). A definition that encompasses this is the following *"evaluation is the careful assessment ex post of the merit of the content, administration, output, outcome effects, and organization of public sector interventions, which is intended to be useful in practical action situations."* (Vedung, 2017).

This means that evaluation is intended to create a basis for practical considerations. Increased institutionalisation would consequently imply greater use of evaluations in government. The question is if this is happening.

From our perspective within the development policy administration, we have paid close attention to the use of thematic and strategic evaluations over the last few years. In our annual report in 2016 (Norad, 2016), when looking at evaluations completed over the last twelve months, findings and conclusions were almost identical with those of evaluations completed in the previous years. A potential explanation for this is that administrations do not learn anything and lack the capacity and culture to bring about change, which indicates limited direct use of evaluations.

The UN study also found that the use of evaluation reports is consistently low for most UN organisations. Even organisations in which the evaluation function is considered to perform well see only an average level of use by the intended audience. Similarly, the OECD study found that the extent to which evaluations are used is mixed.

Another observation related to use is from a recent meta-evaluation of decentralised evaluations in the Norwegian aid administration (Norad, 2017). Although the meta-evaluation found that the decentralised evaluations were well used by the unit responsible for managing the grant for these interventions, it also found that the quality of the decentralised evaluations was unsatisfactory due to weaknesses in methodology and analysis. Over 65 percent of the reviews did not contain sound methodological underpinnings that would support or produce credible findings. Under such circumstances, use may be more harmful than beneficial.

Explanations for the gap

The discussion above points towards the lack of a clear link between the increased institutionalisation of evaluation and perceptions of the use of evaluation. In the following, possible explanations are presented.

One explanation may be the increasingly complex settings that evaluators are dealing with, especially in the field of development evaluation. This complexity relates to the nature of the programmes, the context within which programmes are embedded, the interactions with the many stakeholders, the process of change and causality, and the nature of the evaluation process (Bamberger et al, 2016). Evaluations assessing programmes addressing problems caused by the humanitarian crisis or related to climate change are examples of complexity. This implies that evaluations, no matter how comprehensive, will not be able to deal with all relevant aspects of these topics. Hence, intended users may consider any evaluation too narrow to be used as basis for decision-making.

Another possible explanation is the rhetoric surrounding public policies in general and in particular within the development policy area. In many western countries, this rhetoric is about altruism and poverty. At the same time, many programmes reflect donors' own national interests (Kharas & Rogerson 2017). Examples are programmes to prevent health epidemics spreading to Europe, or to prevent an influx of refugees to the western world. This rhetoric-reality gap may create difficulties for an honest discussion based on evaluation and thus limit the use of evaluation findings and recommendations.

Yet another explanation is the greater demand for reporting and documentation. Taxpayers and politicians expect proof of the results of aid spending. Evaluators are rarely able to supply this sort of documentation. This may undermine the interest in using evaluations. Simple, quantitative reporting may be more attractive than nuanced analytical evaluations.

Closing reflections

If these observations hold true, what can evaluators do? There are three opportunities to close the gap.

The first is to strengthen quality assurance of all evaluation elements to avoid use and misuse of findings that are not credible.

The second is to acknowledge the increased complexity that evaluators are facing and use new evaluation approaches, like evaluation synthesis studies, to draw on a more comprehensive evidence base than what just one evaluation can offer. In addition, a more active and participatory approach to developing recommendations and communicating findings may have a positive effect on use.

The third is to expand an understanding of what use implies. Use may be too narrowly understood as intended use by intended users, i.e. the direct inputting of evaluative findings into immediate decisions. A wider understanding of use may bring a more positive conclusion. Use may also be understood as relating to the evaluation process or as inspiration, enlightenment or to constituting a new understanding (Vedung, 2017).

References

Bamberger, M., Vaessen, J., & Raimondo, E. (Eds.). (2015). *Dealing with complexity in development evaluation: A practical approach*. Sage Publications.

Jacob, S., Speer, S., & Furubo, J. E. (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, *21*(1), 6–31.

Kharas, H., & Rogerson, A. (2017). Global development trends and challenges: Horizon 2025 revisited. *London: Overseas Development Institute*.

Norad Evaluation Department. (2016). *Learning from the past – Towards better development?* Annual Report 2015/16, Oslo: Norad.

Norad Evaluation Department. (2017). *The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation*. Report 1/2017, Oslo: Norad.

OECD. (1991). *DAC Principles for Evaluation of Development Assistance Development Assistance Committee*. Paris: OECD Publishing.

OECD. (2016). *Evaluation Systems in Development Co-operation*. Paris: OECD Publishing.

United Nations. (2014). *Analysis of the Evaluation Function in the United Nations System*. JIU/REP/2014/6. Geneva: Joint Inspection Unit.

Vedung, E. (2017). *Six-times-two Uses of Evaluation*. Uppsala: Evert Vedung's Personal Archive.

■

A BRIEF COMPARISON AND DISCUSSION OF SOME META-EVALUATIONS AND QUALITY REVIEWS OF DEVELOPMENT EVALUATIONS

Ole Winckler Andersen

Several agencies in the field of international development (e.g. bilaterals, multilaterals, non-governmental organizations) have commissioned meta-evaluations and quality reviews of their evaluations. This paper presents a brief comparison of a selected number of these analyses and highlights some issues for consideration in connection with potential future meta-evaluations.

The concept of meta-evaluation is often used to imply "…an evaluation of an evaluation to judge its quality…" (OECD-DAC, 2010).[1] Most quality reviews have a similar focus, and a distinction between meta-evaluations and quality reviews will not be made in the following.

Sample

This paper briefly discusses an assessment of a purposive sample of five meta-evaluations conducted by five different donors in the field of international development (Forss et al., 2008; Perrin, 2009; Hageboeck et al., 2013; Cooney et al., 2015; Chapman et al., 2017).

Comparison

There are various potential approaches to meta-evaluation, but most approaches consider the following four dimensions:

a) The purpose of the meta-evaluation. Does it primarily serve a formative or a summative purpose?
b) The sample of included evaluations. What are the rationale and criteria for the selected sample?
c) The information sources used in the meta-evaluation. Are evaluation reports the only source or are other documents and sources used?
d) The applied criteria in assessing the evaluations. Which criteria are used, and in particular for assessing evaluation quality?

A comparison of the five meta-evaluations shows that all five meta-evaluations, as expected, have a main focus on the quality of Terms of Reference and evaluation reports, while they differ in their coverage of various steps in the evaluation process and the

1 For a full definition, which includes different interpretations of the concept 'meta-evaluation', see (OECD-DAC, 2010).

quality of the evaluation system as a whole. This reflects differences in understanding of evaluation quality as well as of what factors can potentially influence this quality. All five meta-evaluations contain recommendations on how to improve the quality of evaluations, but the primary purpose of the meta-evaluations seems to be accountability. One of the meta-evaluations is also covering the effectiveness of the specific agency's development cooperation based on an analysis of evaluations of good quality in the sample.

The size and selection of the samples of evaluations covered by the individual meta-evaluations vary. The number of evaluations ranges from 14 to 340 evaluations covering various sectors. The low number of evaluations in some of the meta-evaluations prevented more disaggregated analyses; e.g. at sector level. All meta-evaluations include decentralized evaluations,[2] some covering one year and others several years. The selection methods for including individual evaluations in the analysis include random selection as well as purposive selection in consultation with commissioners.

As indicated above, the five meta-evaluations assess the quality of Terms of Reference (TOR) and the evaluation reports, and some of them also review other documents related to the evaluation process. The degree to which the meta-evaluations use other sources of information varies, and different combinations of methods such as interviews, surveys and case studies are used in the review processes. None of the meta-evaluations seems to have involved partner countries.

All five meta-evaluations are based on specific checklists developed by the meta-evaluators. The checklists are inspired by guidelines, such as the OECD-DAC Quality Standards, but adapted to the specific meta-evaluation. Although some overlaps can be found, the content, mixture and number of criteria differ between the five meta-evaluations. It is difficult to compare the number of criteria as the meta-evaluations use different groups and subgroups of criteria, but the number of applied quality criteria is in the range of 30–60. The assumption seems to be that the same standards and criteria can be used for all evaluations under review.

The meta-evaluations contain interesting findings and discussions regarding the relationships between the TOR, the evaluation budget, the composition of the team and the quality of the evaluation report as well as the systemic factors that can potentially influence the quality of various steps in the evaluation process. Examples of general recommendations are that the TOR should be more focused and include improved sections on evaluation methodology, and that efforts should be made to strengthen work on methodological issues in the evaluations.

Discussion

The main purpose of this short paper is to encourage discussion on the use of meta-evaluations of development evaluations. The five meta-evaluations contain a lot of interesting information as well as detailed discussions of five donor agencies' evaluation systems, and the comparison shows that several issues could be considered and discussed in the framework of potential future meta-evaluations.

The five meta-evaluations in the sample serve various purposes with obvious implications for their potential use. The praxis seems to be to apply the same criteria and rating procedures to all evaluations, irrespectively of context, intervention, sector and type of evaluation, where an alternative option would be to have more focused meta-evaluations, e.g. covering a sector or thematic area. In view of that at least two of the agencies are conducting recurrent meta-evaluations, this raises several questions on the role of meta-evaluations in comparison to other quality assurance mechanisms such as independent external peer reviews and quality assurance of ongoing evaluations. A further observation is that the five meta-evaluations have not involved partner countries and therefore do not reflect their potential assessment and use of the conducted evaluations.

The review and comparison of the five meta-evaluations show that no uniform approach exists to meta-evaluations of development evaluations, and that the applied methodological approaches differ. All the meta-evaluations use well-defined assessment criteria, but their number and specification vary as does the sample size of evaluations covered by the studies and sources used for information. These differences obviously reflect the different purposes and contexts of the meta-evaluations, but a more detailed comparison and discussion of the value of the different methodological approaches would be interesting. As indicated, meta-evaluations of development evaluations have not been systematically used, but the five meta-evaluations provide an excellent basis for discussions on the practice and use of potential future meta-evaluations.

References

Chapman, N., Loyd, R., Villanger, E. & Gleed, G. (2017). *The quality of reviews and decentralized evaluations in Norwegian development cooperation*. Oslo: NORAD.

Cooney, G., Rojas, K, Arsenault, M. and Babcock, K. (2015). Meta-evaluation of project and programme evaluations in 2012–2014. Ministry for Foreign Affairs of Finland.

Forss, K., Vedung, E. Kruse, S. E. Mwaiselage, A. and Nilsdotter (2008). Are Sida Evaluations Good Enough? Sida Studies in Evaluation 2008:1.

Hageboeck, M., Frumkin, M. and Monschein, S. (2013). Meta-evaluation of the quality and coverage of USAID Evaluations 2009–2012. USAID, August 2013.

OECD-DAC (2010). Glossary of Key Terms in Evaluation and Results Based Management. OECD Publications, Paris.

Perrin, B. (2009). Review of the Quality of DFID Evaluations. IACDI. The Independent Advisory Committee on Development Impact, September 2009.

■

2 Evaluations that are not commissioned/managed/conducted by an (independent) centralized evaluation unit but by operational units.

THE AUTHORS

Ole Winckler Andersen

Ole Winckler Andersen is Senior Analyst at the Danish Institute for International Studies, where he is focusing on development finance and evaluation. He holds a M.Sc. in Economics and a Ph.D. in Public Administration and has worked for the Danish Ministry of Foreign Affairs for more than 20 years, including as Head of Evaluation Department (2017–13) and Deputy Permanent Representative to the OECD (2013–17). Before that he was assistant and associate professor at Roskilde University in Denmark. He has managed a number of evaluations and has been member of management committees for several international development evaluations. He has published on evaluation in various international journals. He has since 2015 been a member of the Board of EES.

Per Øyvind Bastøe

Per Oyvind Bastoe is the Evaluation Director at Norwegian Agency for Development Cooperation and the chair of the OECD/DAC Evaluation Network. He has broad experience from international development and evaluation and has previously held senior positions in other parts of the Norwegian government administration, in the World Bank, the Asian Development Bank and the Inter-American Development Bank. He is a member of the International Evaluation Research Group and has published several books and articles on development policy, evaluation and organizational change.

Jan Hense

Jan Hense is full professor for Higher Education Learning and Evaluation at the Justus-Liebig-University Giessen (Germany) and president of the German and Austrian *Gesellschaft für Evaluation*. He holds a PhD in educational psychology with an emphasis in evaluation (2005) from the Ludwig-Maximilians-University in Munich (Germany). His research areas include evaluation theory and research on evaluation, innovative teaching and learning methods, and technology based learning. As an evaluator and evaluation consultant he has been involved in many local, national and international evaluation projects. Within DeGEval he contributed to the recent revision of DeGEval's evaluation standards in a leading role, and has been involved in several other initiatives related to the professionalization of evaluation.

Karen Jørgensen

As Head of the Review, Evaluation and Engagement Division of the OECD Development Cooperation Directorate for more than a decade, Karen Jorgensen led more than 60 reviews of DAC members' development cooperation programmes and in addition oversaw the work of the Development Evaluators' Network (Evalnet). Prior to that she served as UNDP's head of programme in Afghanistan from 2003–2006, as Assistant Director for the Sustainable Energy and Environment Division and head of division in the Audit office of UNDP. Ms Jorgensen is a Norwegian national.

Christiane Kerlen

Dr Christiane Kerlen, director at Kerlen Evaluation Ltd, is an independent evaluator and evaluation consultant with over 15 years of experience. Her key areas of expertise include evaluation design and development of tailored evaluation instruments, predominantly for research, technology and innovation (RTI) programmes as well as programmes and projects in foundations and for institutional evaluations. Since 2015 Christiane is vice-president of DeGEval – Gesellschaft für Evaluation e.V. having been a member of DeGEval's board since 2011.

Siv J. Lillestøl

Siv J. Lillestøl is Assistant Director in the Evaluation Department in the Norwegian Agency for Development Cooperation (Norad). Prior to joining the department in 2009 she worked several years with monitoring and evaluation in the The Joint United Nations Programme on HIV/AIDS (UNAIDS) and in the World Food Programme (WFP). She has served at the governing boards of the Norwegian Evaluation Society and the Network for evaluation in the Norwegian public administration.

Wolfgang Meyer

Wolfgang Meyer (Germany), PD Dr., Assistant Professor at Saarland University, Saarbruecken, and Adjunct Professor at Uganda Technology and Management University, Kampala; Vice-Director of the Centre for Evaluation; Founding Member of the German Evaluation Society; Member of the European Evaluation Society; Sociologist with a Focus on Empirical Methodology; Evaluation Specialist in the Field of Development Cooperation, Environment, Labor Market and Regional Development.

Reinhard Stockmann

Dr. Stockmann is Professor of Sociology at Saarland University, the founder and director of the Center for Evaluation (CEval), the Managing Director of the Master-Course of Evaluation at Saarland University and of the first European study program "Master of Evaluation" in a blended learning format. He is also the Managing Publisher of the German Zeitschrift für Evaluation (Journal for Evaluation) and a founding member of the German Society for Evaluation (DeGEval).

Kevin Williams

Kevin Williams is head of evaluation at the Organisation for Economic Co-operation and Development (OECD). He has extensive experience of building evaluation capacity and embedding evaluation practice within organisations, particular at international and supranational levels. He has set up evaluation functions in both the OECD and the European Commission, and has contributed to the development of practical evaluation guidance and training initiatives in a range of policy areas including EU regional development programmes, economic policy and the Common Agricultural Policy. He has also contributed towards the development of frameworks to adapt evaluation practice to a range of non-programmatic evaluand, including intergovernmental committees and legislative initiatives in both ex ante and ex post contexts. In his current role he is responsible for the internal evaluation of OECD committees and programmes. He graduated in European economics via the Erasmus programme from the universities of Nantes and Middlesex and holds a master's degree in industrial economics from Université Lumière Lyon II.