



WORKING PAPER MARCH 2021
RELEVANT? ALMOST ALWAYS
RELEVANCE IN DEVELOPMENT COOPERATION

Joel Samoff

Relevant? Almost Always – Relevance in Development Cooperation

Joel Samoff

Underlagsrapport 2021
till
Expertgruppen för biståndsanalys (EBA)

Joel Samoff

An educator, researcher, and evaluator, Joel Samoff combines the scholar's critical approach and extensive experience in international development. He studies the links among research, public policy, and foreign aid. At Stanford University since 1980, he holds honorary doctorates from the University of Pretoria and the University of the Free State in South Africa. Recent publications include *Capturing Complexity and Context: Evaluating Aid to Education* (with Jane Leer and Michelle Reddy) and *Higher Education for Self-Reliance: Tanzania and Africa*.

Please refer to the present working paper as: EBA Joel Samoff (2021) *Relevant? Almost Always – Relevance in Development Cooperation*, EBA Working Paper March 2021, Expertgruppen för biståndsanalys.

This working paper can be downloaded free of charge at www.eba.se

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Printed by Elanders Sverige AB

Stockholm 2021

Cover design by Julia Demchenko

Table of Contents

Foreword by the EBA	1
Sammanfattning	3
Summary	7
Relevant?	10
Approaches to Evaluating Development Assistance.....	11
Relevance: OECD Evaluation Criterion	12
Relevance: Doing the Right Things	13
Guidance on Implementing the Relevance Criterion	15
Intervention.....	16
Relevance: Sida Additional Specifications	18
Applying Sida’s Evaluation Criteria: Assessment and Learning.....	19
Relevant: Almost Always	20
Relevance as a Gatekeeper	22
Relevance: Impossible to Assess	23
Sida’s Relevance Criterion	24
Relevant by Definition	25
Relevant by Evaluator’s Choice	27
Explaining the Selected Referent: Complicating Challenges	28
Relevance is not a Fixed Target	31
Relevance is necessarily a relational criterion.....	31
Relevance is always time constrained.....	32
Assessing relevance always requires attention to who makes the assessment	33
Relevance Assessments: Observations.....	34

Relevance: Remove, Revise, Reinvigorate	36
Remove	36
Revise	37
Reinvigorate.....	43
References.....	45

Foreword by the EBA

The use of the OECD/DAC evaluation criteria (relevance, coherence, effectiveness, efficiency, impact and sustainability) is now standard in most evaluations of development cooperation activities.

The present study focuses on the relevance criterion. An apparent problem with the use of this criterion is that evaluations almost always find that interventions in development cooperation are relevant.

Why on earth would such a finding be problematic? Don't we want interventions to be relevant? Sure. However, learning requires variation in outcomes, and the value of a signal always providing the same type of information – that something works – is limited, and spending time and effort on narratives to already known answers is not an efficient use of evaluation budgets. In short, if interventions are always relevant, evaluating relevance lacks relevance.

In practice, evaluating relevance is challenging. It is important, therefore, to understand better the relevance criterion and its use in evaluation. Several dimensions of this issue require attention, such as how relevance is understood in development cooperation and how those different understandings and priorities are incorporated in evaluations.

Against this background, the EBA asked Joel Samoff to explore the relevance criterion from three perspectives: (i) How has the evaluation criterion relevance been assessed in previous (selected) Sida evaluations? (ii) What types of information and analysis are necessary to determine if a particular activity has been relevant? (iii) How might current practice in evaluating relevance be improved to make the criterion more useful?

The result is a thought-provoking conversation starter summed up in nine recommendations. Our intention is that the report will be of relevance to commissioners of evaluations at Sida and elsewhere as well as evaluators of development interventions, and to contribute to a more nuanced use of the criterion. Thus, our hope is that it will contribute to better evaluations, in companion with guides such as OECD/DAC's "Thoughtful evaluation" and Sida's Evaluation handbook.

The EBA Working Paper Series constitutes shorter overviews, surveys, mappings and analyses that have been undertaken to bring about discussion and advance knowledge of a particular topic. Working Papers are not subject to any formal approval process by the Expert Group. Just as in the EBA reports, authors are solely responsible for the content, conclusions and recommendations.

Stockholm, March 2021

Jan Pettersson

Sammanfattning

Utvärderingar av det svenska biståndet kommer nästan alltid fram till att insatserna är relevanta. Förutom att ge en anledning att fira framgången så har resultatet dock begränsad nytta och bidrar varken till Sidas resultatbaserade styrning eller till att förbättra genomförande i Sidas partnerorganisationer. Detta motiverar en ny granskning och ett kritisk utforskande av relevanskriteriets användning i utvärderingar av utvecklingssamarbetet. Grunden för granskningen inkluderar nyligen utförda utvärderingar gjorda av och för Sida, ett urval av utvärderingar från andra internationella organisationer, utvärderingsvägledning från OECD och Sida samt forskning om utvärdering.

Undersökningen av relevanskriteriet tar sin början i Sidas definition av relevans:

I vilken utsträckning interventionens mål och utformning svarar mot: mottagarnas behov på global, nationell och partner/institutionsnivå, policyer och prioriteringar, och fortsätter att göra detta om omständigheterna förändras. (EBAs översättning)

Definitionen skapar i sig flera utmaningar för tillämpningen av kriteriet. För det första, Sida tolkar relevans som "att göra rätt saker." Är det då utvärderarens roll att avgöra om aktiviteten är positivt relaterad till något (dvs. relevant) eller om insatsen är korrekt (gör rätt saker)? Vem som ska bestämma vad som är rätt lämnas obesvarat, liksom tillvägagångssättet när det finns olika föreställningar om vad som är rätt. Ett andra problem är antagandet om att olika behov, policyer och prioriteringar är kompletterande och förenliga. I praktiken kolliderar de ofta. Kriteriet och dess guide nämner inte dessa spänningar och ger inte tillräcklig vägledning för hur de ska inkluderas i bedömningen. För det tredje så är det inte möjligt att adressera alla dimensioner som anges i relevanskriteriet, även för den mest omfattande utvärderingen. Utan ytterligare vägledning så blir det för brett, för komplext och för motsägelsefullt att bedöma insatsers relevans.

Det faktum att relevanskriteriet fungerar som en ”gatekeeper” - om en aktivitet inte är relevant, varför undersöka om fel sak har gjorts effektivt? – uppmuntrar till snabba, formalistiska och ytliga bedömningar. Hur har utvärderarna då agerat?

För det första genom förenkling. Utvärderarens utgångspunkt är Sidas antagande om relevans. Om det inte finns övertygande bevis för det motsatta, ska relevansen bekräftas och insatsen blir då relevant per definition. Alternativt så är relevansen utvärderarens val. Relevansen kan då bekräftas genom att aktiviteten överensstämmer med en formell policy, ett officiellt mål eller prioritering, eller ett av de globala hållbarhetsmålen eller ett bredare område, som jämställdhet. Då många sådana referenspunkter är både trovärdiga och försvarbara behövs ingen omfattande diskussion eller motivering.

Även om varje enskild bekräftelse av relevansen verkar rimlig så blir det kollektiva resultatet en kakofoni och inte ett samstämmigt resultat.

Hur kan relevanskriteriet då bli mer användbart?

Utvärderare står inför flera utmaningar för att bedöma relevans. För det första så är relevans ovillkorligen ett kriterium som används i relation till något. Relevant för vad? Och för det andra, i praktiken blir «relevant för vad?» alltid «relevant för vem?». Vems behov, policyer och prioriteringar bör ha högsta prioritet? Varför deras? Den tredje utmaningen är klyftan mellan formella dokument och praxis. Kan relevans exempelvis bekräftas genom överensstämmelse med en uttalad policy som ignoreras och har begränsad praktisk användning? Än mer problematisk är den fjärde utmaningen. Som tidigare beskrivits så antas här att det finns en samstämmighet mellan behov, policyer och prioriteringar och med angivna referenspunkter. Här finns dock en uppsjö av motsättningar. Detta när behov, politik och prioriteringar är i konflikt, där en utvecklingspartner och Sverige har olika policyer och olika prioriteringar, och där en viss aktivitet och dess institutioner är överhopade av spänningar, vad är relevant? För det femte så finns det ofta en glidning mellan relevans och andra utvärderingskriterier. Utvärderare använder en aktivitets måluppfyllelse eller effekt som bevis för relevans. För det sjätte, eftersom relevansen alltid är tidsbegränsad, måste en användbar bedömning tidsbestämmas. För det sjunde, för att en utvärderares

professionella bedömning av relevans ska vara mer än en ”tick-box”, så måste den vila på systematisk insamling, analys och sammanställning av data. En åttonde utmaning ligger i att bestämma vems perspektiv som ska ha störst vikt i utvärderingsprocessen. Om inte frågan "vems bedömning av relevans är detta?" besvaras så begränsas användbarheten av relevanskriteriet, det försvagar utvärderingen och begränsar dess möjlighet att bidra till utformningen och genomförandet av utvecklingspolicy. För det nionde så är vägledningen i processen – från OECD-DAC och Sidas kriterier och handböcker till uppdragsbeskrivningar för utvärderare av utvecklingsaktiviteter, till de utvärderingar de producerar – ojämn och ofta ofullständig.

Vilken framtida roll har då relevanskriteriet i Sidas decentraliserade och partnerledda utvärderingar? Sluta använda? Revidera?

Sluta använda kriteriet. Eftersom själva beslutet att tillhandahålla bistånd i sig kräver en bekräftelse av att biståndet är relevant, eftersom relevans är ett kriterium med många svårfångade utmaningar, och flertalet utvärderare rapporterar att vad de än utvärderar så är det relevant så kan relevanskriteriet tas bort i de flesta utvärderingar. Sida kan istället identifiera några särskilt utvalda utvärderingar där relevansen ska bedömas.

Revidera kriteriet. Alternativt kan Sida behålla relevanskriteriet genom att revidera dess innehåll och användning för att göra det mer informativt och bidragande. En hög prioritet är att ta itu med de noterade problemen, inklusive att skilja mellan relevans (vad som är relaterat och lämpligt) och "att göra rätt saker" (vad som bör göras). Att istället bedöma relevans längs ett kontinuum, snarare än en dikotomi, kräver ett systematiskt och kritiskt fokus på evidens. Att erkänna att relevansen regelbundet är omtvistad kräver att utvärderare utforskar alternativa bedömningar. För att utveckla perspektiv och insikter hos de som är involverade i biståndsaktiviteter kan fördelarna med deltagande utvärdering visa sig vara viktigare än dessas begränsningar. Utvärderingsinstitutioner i partnerländer, med stöd från Sida i de fall det behövs, bör ha det primära ansvaret för att utvärdera en aktivitetens relevans.

För att stärka relevanskriteriet kommer det också krävas att Sida efterfrågar en oberoende, välgrundad och kritisk bedömning av relevansen som tillåts dra slutsatsen att finansierade aktiviteter strider mot Sidas eller dess partners policy, behov och prioriteringar.

Att bedöma relevans är samtidigt viktigt. Genom att kritiskt analysera kriteriets beståndsdelar blir bedömningen också möjlig. En omfördelning av ansvaret för granskningen kan här både underlätta resultatbaserad styrning och fördjupa och stärka partners deltagande i utvärderingsprocessen.

Summary

Recent evaluations have almost always found Swedish development assistance relevant. Yet, beyond celebrating success, that common finding has limited utility, supporting neither Sida's results-based management nor partners' improved implementation. That warrants a fresh look and a critical exploration of the relevance criterion in evaluations of development cooperation, its use, and its assessment. The foundation for that exploration includes recent evaluations undertaken by and for Sida, sample evaluations for other agencies, OECD and Sida evaluation guidance, and evaluation research.

The review of the relevance criterion begins with Sida's specification:

“The extent to which the intervention objectives and design respond to beneficiaries' global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.”

That specification itself creates several immediate application challenges. First, Sida, understands relevance as “doing the right things.” Are evaluators to determine whether evaluated activities are positively related (relevant) or correct (doing the right things)? Who determines what is right is left unaddressed, as is how to proceed when different notions of what is right diverge. A second problem is the assumption that diverse needs, policies, and priorities are complementary and consistent. In practice, often they clash. The criterion and its associated guidance do not recognize those tensions or provide adequate direction for evaluators on incorporating them in their assessments. Third, addressing all of the dimensions specified in the relevance criterion is beyond the reach of even the most extensive evaluation. In the absence of further guidance, relevance is too vast, too complex, and too conflicted to assess.

That the relevance criterion is a gatekeeper – if an activity is not relevant, why examine whether the wrong thing has been done efficiently? – encourages rapid, formulaic and superficial assessments. How have evaluators responded?

First, simplify. Evaluators' starting point is Sida's presumption of relevance. Unless there is compelling evidence to the contrary, relevance is to be confirmed – relevant by definition. Alternatively, relevant by evaluator's choice. Relevance can be confirmed by consistency with a formally stated policy, or other official indication of goals, objectives, or priorities, or a Sustainable Development Goal, or a broad objective, say gender equity. Since many referents are plausible and defensible, none needs extensive discussion or justification.

While each confirmation of relevance seems reasonable in its own terms, the collective result is cacophony, not coordination. How might the relevance criterion become more useful?

Evaluators face several additional implementation challenges in assessing relevance. First, relevance is necessarily a relational criterion. Relevant to what? Second, in practice, «relevant to what?» always becomes «relevant to whom?». Whose needs, policies, and priorities should have the highest priority? Why theirs? Third is the gap between formal documents and actual practice. Can relevance be confirmed by congruence with a stated policy that is widely ignored and has little practical consequence? Even more problematic is a fourth challenge. As written, the relevance criterion assumes a general convergence among needs, policies, and priorities and then across the specified reference points. Yet, conflicts among them abound. Where needs, policies, and priorities are in tension, where a development partner and Sweden have divergent policies and different priorities, and where a particular activity and its institutions are buffeted by those tensions, what is relevant? Fifth, there is frequent slippage between relevance and other evaluation criteria. Evaluators use an activity's effectiveness or impact as evidence of relevance. Sixth, since relevance is always time constrained, a useful assessment of relevance must be time-stamped. Seventh, for an evaluator's professional judgment on relevance to be more than a tick box, it must rest on the systematic collection, analysis, and presentation of supporting evidence. An eighth assessment challenge is determining whose perspective is to have greater weight in the evaluation process. Failure to address «whose assessment of relevance is this?» limits the utility of the relevance criterion, weakens the overall evaluation, and constrains its ability to contribute to the formulation and implementation of development policy. Ninth, the guidance path – from OECD-DAC and Sida criteria and guidance to

terms of reference for evaluators of development activities to the evaluations they produce – is rocky and regularly incomplete.

What role, then, for the relevance criterion in Sida’s decentralized and partner-led evaluations? Remove? Revise?

Remove. Since the decision to provide development assistance requires affirming its relevance, since relevance is a troubled criterion that is difficult to capture, and since most evaluators report that whatever they are evaluating is relevant, the relevance criterion can be removed as a high priority focus for most post-activity evaluations. Sida might then identify selected evaluations in which relevance is to be assessed.

Revise. Alternatively, Sida can retain the relevance criterion, revising its content and implementation to make it more informative and productive. High priority is addressing the problems noted, including distinguishing between relevance (what is related and appropriate) and “doing the right things” (what should be done). Assessing relevance along a continuum, rather than a dichotomy, requires systematic and critical attention to evidence. Recognizing that relevance is regularly contested requires evaluators to explore alternative assessments. To develop the perspectives and insights of those involved in supported activities, the advantages of participatory evaluations can prove far more consequential than their limitations. Partner country evaluation institutions, as needed with Sida support, should have primary responsibility for post-activity evaluations of relevance.

Reinvigorating the relevance criterion will require communicating that Sida seeks an independent, well-grounded, and critical assessment of relevance, which may include the report that funded activities are at odds with Sida’s or its partners policies and priorities.

Assessing relevance remains important. Unpacking the relevance criterion makes that possible. Reorganizing responsibility for its assessment can both facilitate results-based management and deepen and strengthen partners’ participation in the evaluation process.

Relevant?

Recent evaluations have almost always found Swedish development assistance relevant. Beyond celebrating success, that common finding has limited utility, supporting neither Sida's results-based management nor partners' improved implementation.

All organizations involved in development cooperation intend the activities they support be relevant. That is understood to mean that supported activities should clearly contribute to development partners' priority goals and objectives. Ideally, supported activities within partner countries should also contribute to established external goals, both global (for example, the Sustainable Development Goals) and specific to the funding source (for example, gender equity or democratic practice).

With those strong expectations, funders regularly expect evaluators of development assistance to assess relevance. Where evaluators do assess relevance in Swedish development cooperation, it turns out, nearly all funded activities are found to be relevant. Is the activity relevant, the evaluators ask. Certainly! they report. That may reflect Sida's discerning decisions about what is to be funded and Sida's partners' outstandingly effective implementation. Or that near universal report that funded activities are relevant may reflect problems with the relevance criterion. Even where the success rate is very high, an evaluation criterion with no significant variation contributes little to learning from experience.

From a somewhat different perspective, we can observe that everyone involved in development cooperation regards relevance as important. Everyone is confident about the criterion – they know what relevance means and how to achieve it. Yet, on examination, we find that understandings of relevance diverge sharply. The assumption of a common meaning regularly obscures fractured practice.

That nearly all development assistance is assessed to be relevant, notwithstanding different and sometimes incompatible notions of what relevance means, warrants a critical exploration of the relevance criterion, its use, and its assessment. Accordingly, my concern here is to develop an overview of the issues and suggestions for making relevance a more useful evaluation criterion. For that, I draw on significant earlier systematic and

insightful analyses. I have as well reviewed evaluation guidance developed by several organizations (including OECD-DAC, Sida, World Bank, United Nations Evaluation Group, and USAID), recent summary and detailed evaluations (primarily Sida, with samples from Germany, Netherlands, World Bank, USAID, Japan), and the academic research on evaluations.

Exploring the relevance criterion invites attention to evaluation more broadly, especially to the roles and uses of evaluations, and to foreign aid. To focus concentrated attention on relevance, however, those and related topics must be beyond the scope of this review.

Approaches to Evaluating Development Assistance

Relevance is not a universally applied evaluation criterion.

A review of recent evaluations of several organizations' development assistance, both national (for example, Sweden, Germany, United States) and multi-national (OECD, United Nations, World Bank), highlights an important difference in starting point. Some evaluations use the stated objectives of a particular supported activity to develop evaluation indicators and to assess the activity and the assistance to it. For example, USAID supported an initiative in Ethiopia to improve social services for highly vulnerable children. The evaluators used the support's four stated objectives to craft questions to organize the evaluation. For example, to what extent has participation in the programme strengthened regional and local government capacity to respond to the needs of highly vulnerable children and their families? And, what have been the changes in regional and local government services that can be attributed to the supported programme? The evaluation begins and remains within the stated objectives. Unless they are specified in the support objectives, attention to, say, the global Sustainable Development Goals, or to human rights, is neither expected nor addressed. Where funding agencies expect evaluations to be focused on stated assistance objectives with no explicit instruction to explore other criteria, evaluations generally do not examine or assess relevance directly.

Other evaluations address a set of concerns that are developed outside a particular supported activity and its evaluation, for example, achieving gender equity or one of the Sustainable Development Goals, that may be referenced externally or embedded within the assistance arrangement. Relevance lies here. Sweden considers relevance a high priority criterion for assessing its development assistance across many different sorts of activity. Though there is some flexibility in the selection of criteria, generally, evaluators are expected to assess relevance whether they are evaluating support to forestry management, early childhood education curriculum, or local government capacity building.

The discussion here explores relevance as an expected criterion for evaluating development assistance, whether or not it is explicitly embedded in the objectives specified for an assistance arrangement.

Next, it is important to consider the specification of relevance in development assistance, following a path that leads from OECD to Sida and that highlights several of the relevance criterion's problematic dimensions.

Relevance: OECD Evaluation Criterion

A primary point of reference for defining and applying relevance as a criterion in the evaluation of development assistance has been the work of the Development Assistance Committee (DAC) of the Organisation for Economic Cooperation and Development (OECD). Originally laid out in 1991, OECD-DAC's evaluation criteria have been revised and updated, most recently in the context of the United Nations' adoption of Sustainable Development Goals with a 2030 target (OECD 2019). Each of the six criteria (relevance, coherence, effectiveness, efficiency, impact, and sustainability) is explained conceptually, with details and guidance on application.

OECD-DAC is clear that the evaluation criteria “play a normative role,” describing the “desired attributes of interventions” (OECD 2019: 5). That is, beyond outlining how to assess what has happened, the evaluation criteria are to specify what should happen. In that role, they function to (1) tell evaluators how to frame their assessments, (2) tell the managers of aid-supported activities how their work will be reviewed and therefore

what should receive priority attention, and (3) signal to aid recipients the expectations of the aid providers.

The OECD-DAC criteria reflect many years of discussion by many people across many agencies. Their periodic revisions reflect the evolution of development thinking over time and world circumstances as well as compromises among the participating countries' primary and particular concerns. Regularly, evaluators have identified evaluation problems and sought to address them. Reviewing the contours and depth of that process is beyond the scope of this brief analysis (Chianca 2005). My task here is a fresh look at the relevance criterion, prompted by the puzzlement and frustration that nearly all evaluated activities are found to be relevant.

Relevance: Doing the Right Things

The OECD-DAC specification of the relevance criterion both gives it more weight and complicates its assessment. For OECD-DAC, relevance is “doing the right things.”

I am unclear on the history of that association, but in everyday language in English, it is confusing. Relevance has to do with being related or connected, and sometimes appropriately connected. “Doing the right things” refers to someone’s judgment on what is correct, or what should be done. If the intent is to evaluate connectedness, then the equation with “doing the right things” gets in the way. If the intent is to evaluate whether or not Sida or another organization is doing what it should be doing – funding activities expected to make clear progress toward one or another of Sida’s high priority objectives – then relevance gets in the way. For the discussion that follows I assume that the intent is both relevance and doing the right things.

The equation of those two ideas converts the common language notion of relevant – related; appropriate; consistent with a larger objective – into an imperative. In everyday discourse, doing something that is inappropriate or less appropriate may be a serious problem, but doing something that is wrong is far worse. While there may be a strong case for pursuing a course of action that departs from other evaluation criteria, say, is less efficient or that will be difficult to sustain, there can be no case for doing the wrong thing. Support to an inclusive strategy for curriculum development, for

example, may be deemed appropriate, even if an alternative approach can be expected to cost less. But training in use of computers for teachers who are assigned to schools with no electricity is not “doing the right thing,” even if the training is cost-effective and consonant with a national policy to improve teacher education.

The specification of the relevance criterion creates additional evaluation challenges. For OECD-DAC, relevance (doing the right things):

“The extent to which the intervention objectives and design respond to beneficiaries’ global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.”

The attention to beneficiaries – those affected by the supported activity – is intended to shift the emphasis from relevance to the aid provider’s objectives to relevance to the aid recipient’s needs and interests. An accompanying note explains whose needs, policies, and priorities are to be considered in assessing relevance: “Partner/institution” includes government (national, regional, local), civil society organisations, private entities and international bodies involved in funding, implementing and/or overseeing the intervention.”

Note the assumption that diverse needs, policies, and priorities are complementary and consistent. In practice, often they clash. Most problematic for assessment is the absence of explicit recognition that the specification includes orientations that are likely to be in tension and therefore the absence of guidance on how to address those tensions. Seeking to be relevant to the local needs of a particular activity (say, developing instructional materials readily usable by teachers) may be in tension with relevance to national needs and policies (say, resource availability or national language policy) and to partner’s policies and priorities (say, gender equity or democratic practice). The assumption of a harmonious convergence of multiple interests neither reflects the actual experience of development assistance nor contributes to critical and demanding evaluation. That those who commission evaluations must wrestle with these tensions is desirable, but they and their evaluators would be assisted by a clear recognition that the problem exists.

For OECD-DAC, relevance is doing the right things. But who determines what is right is left unaddressed, as is how to proceed when different notions of what is right clash.

Guidance on Implementing the Relevance Criterion

OECD-DAC has designed its evaluation criteria to be broadly applicable in many situations and offers situational guidance for their application. OECD-DAC cautions that the evaluation criteria must be sensitive to the context of the activity being evaluated and its stakeholders and that the criteria must not be applied mechanistically. As well, the application of the general criteria depends on the purpose of the evaluation.

That concern with context and the ambiguities embedded in the specification suggests that each evaluation must consider tensions among needs, policies, and priorities and must determine whose needs, policies, and priorities are to be the reference point for assessing relevance. In practice, that means that those who commission the evaluation must provide guidance on how relevance is to be assessed.

As well, that concern with context and ambiguities suggests that the assessment of relevance should be considered “to what extent?” and “in what ways?” questions rather than a “relevant/not relevant?” determination.

Also problematic is that attention to relevance regularly slips into other criteria used to assess development assistance. Asking “is this activity relevant?” frequently becomes “is this activity effective?” or “what is its impact?” Those conflation of meaning are not surprising. If an activity is not effective or has no impact, how can it be relevant? With that in mind, it is useful to recall OECD-DAC’s evaluation criteria and their brief explanations:

- | | |
|----------------|---|
| Relevance: | Is the activity doing the right things? |
| Coherence: | How well does the activity fit? |
| Effectiveness: | Is the activity achieving its objectives? |

Efficiency:	How well are resources being used?
Impact:	What difference does the activity make?
Sustainability:	Will the benefits last?

OECD-DAC has periodically offered guidance on the use of these criteria and is, I understand, preparing a new usage guide. That guidance is informed by experience and emphasizes a context-sensitive use of the evaluation criteria. Even where that guidance anticipates some of the problems outlined in this review, those problems persist. There is, it seems, a rocky path, first from the OECD-DAC criteria and guidance to terms of reference for evaluators of development activities, and then from their terms of reference to the evaluations they produce. Addressed in the guidance, ambiguities and tensions in the relevance criterion have largely been sidelined or resolved in the formal expectations for evaluators and in their interpretation of their task.

Intervention

Most OECD-DAC documents, as well as those of the funding and technical assistance agencies, use the term “intervention” to refer to what is to be evaluated. That term has solid roots in the notion that the highest quality social science research, with evaluation as a specialized sort of research, takes the form of a quasi-experiment. The researcher [evaluator] explores the consequences of changes in an independent variable for a dependent variable of interest. Modifying the independent variable is termed an intervention. Ideally, all other influences are to be held constant or treated in the research design so that they do not bias outcomes. For example, introducing a new approach to teaching primary school mathematics can be assessed by measuring mastery outcomes. Or the effectiveness of targeted reorganization of roles and responsibilities in the agriculture ministry might be assessed by measuring the use of extension services.

The assumption of the quasi-experiment as the appropriate model for social science research in general and specifically for evaluations is subtle, often not explicitly recognized, and embedded in the language commonly used. “Intervention” becomes a synonym for activity, or program, or

reform, without critical attention to the perspective and positioning that “intervention” carries with it.

Yet, “intervention” is a particularly inappropriate characterization of most development activities. Research and experience are very clear that the most effective and sustainable development initiatives are those that are owned by the people who must implement them. That ownership is most likely and strongest when those involved play a critical role in conceiving and developing the activity. Those locally rooted roles are not interventions. Interventions are what outsiders do. When a teacher tries a new way of organizing learner work groups, she does not regard her initiative as an intervention. Its organic roots enable the teacher to see her own imagination in and responsibility for the new scheme.

“Intervention” moves us to the perspective of the funding agency. Subtly, and for many who use the term, unintentionally, that devalues the perspective of participants in the aid-supported activity. Intervention is something done to or for aid recipients. The terminological challenge is to find words that centre aid recipients’ imagination and initiative.

In development activities, “intervention” emerges from a medical metaphor. From the perspective of the aid provider, the aid recipient, or one of its sectors or sub-sectors, is not doing well and needs help. The aid provider then becomes the expert diagnostician, who, perhaps working with the aid recipient but often not, determines what is wrong and why and then what are the appropriate remedies. A narrow notion of objectivity requires that the diagnostician remain detached from the patient and limits the patient’s participation in both diagnosis and remedy. After diagnosis the aid provider can supply remedies deemed appropriate, sometimes with attached conditions to encourage the aid recipient to take the bitter medicine in the appropriate doses and times. Though fundamentally problematic and though often disavowed, that medical metaphor remains common in development assistance. It is also sharply at odds with the oft-stated commitment that development assistance should follow national goals and priorities.

More generally, the quasi-experiment is the appropriate model for development research and evaluation only in very limited circumstances. Frequently in the wild – outside laboratories – causes are part of effects

and effects part of causes, other influences must be encouraged, not frozen, and context and complexity are more important than sharply defined and exclusive independent and dependent variables.

Pursuing these issues further is far beyond the scope of this paper. I have explored them briefly here because “intervention” is so widely and so uncritically used and because, most important, words shape our thinking. Here, I refer simply to development activities.

Relevance: Sida Additional Specifications

Sida adopts and extends the OECD-DAC evaluation criteria. As an active participant in the Evaluation Network of the OECD Development Assistance Committee, Sida has contributed to the development, updating, and refinement of the OECD-DAC evaluation criteria. Evaluations at Sida, both the larger and smaller scale evaluations directly commissioned by Sida and those undertaken by Sweden’s development partners, sometimes with Sida funding, are to begin with those criteria. As well, Sida’s evaluation handbook recognizes the importance of evaluation context and purpose.

“All criteria shall be considered for every evaluation, but Sida does not require that all evaluation criteria are included in each evaluation. Rather, it is advisable to focus on a few evaluation criteria, guided by the intended use of the evaluation.” (*Sida’s Evaluation Handbook*: p. 8)

Indeed, Sida’s regular reviews of the evaluation process have been reflective and self-critical and have anticipated some of the issues discussed here. Sida’s Evaluation Handbook is a model of content and clarity and sets a high standard for evaluations. As for the OECD-DAC evaluation guidance, clarity and insights are lost between Handbook and evaluation terms of reference, and between evaluation terms of reference and evaluations.

From that foundation, my task here is to focus fresh attention on a single evaluation criterion, relevance.

Applying Sida's Evaluation Criteria: Assessment and Learning

Sida's evaluations are to be utilization-focused, that is identifying the intended users of an evaluation and then designing it to address their needs and interest. Regarding evaluation as an opportunity for learning and for the development of partnership, Sida expects aid recipients to be involved in the development and implementation of evaluations. Ideally, that orientation should increase the attention to the relevance of the evaluated activity.

A systematic review of a large set of evaluations found that evaluation practice does not meet these aspirations (Samoff, Leer, Reddy 2016). With rare exceptions, evaluations of development assistance are oriented toward the funds provider. Only infrequently do they function as learning opportunities. Most often, they remain unread.

Let us return to the criteria for evaluation. Supplementing the six basic evaluation criteria, four additional criteria are to be addressed in evaluations of humanitarian assistance: connectedness (which replaces the sustainability criterion), coverage, coherence, and coordination.

Beyond those criteria, Sida's own high priority objectives are also to be assessed: contribution to poverty reduction, implementation consistent with a poor people's perspective and a human rights-based approach, conflict-sensitive project design, support for gender equality, and contribution to protecting the environment.

Thus, though first on the list, relevance is one among 14 objectives that are to be evaluated, with a context-sensitive determination of which objectives should receive attention. The length and complexity of that list generate three important risks. One is an evaluation spread thin: many dimensions to be assessed, each with limited attention. The second is an evaluation chaotically executed: the evaluator selects dimensions to assess with no clear plan and perhaps no correspondence with the intent of the activity's initiators or the priorities of the activity's participants. The third is inattention to the ways in which the criteria themselves are in tension with each other.

In its Evaluation Handbook, Sida's examples of questions that can be used to assess relevance are in the form of "to what extent . . . ?" That is, Sida expects the evaluation of an activity's relevance to go beyond the determination of relevant/not relevant. At the least, the evaluation should report more or less relevant. Well executed, an evaluation should explain that assessment, perhaps with examples of specific activities. Here we see the gap between Sida's guidance and the evaluation reports. In the evaluations reviewed, evaluators have generally addressed relevance as a dichotomy – is the activity relevant or not – rather than as a continuum.

As well, although the evaluation criteria are intended to be distinct, often the distinctions are blurred. For example, Sida's second example for assessing relevance reflects the frequent slippage between larger scale objectives – relevance (doing the right things) – and how things are done, captured in the effectiveness (achieving objectives) and efficiency (using resources well) criteria:

“To what extent [have] lessons learned from what works well and less well been used to improve and adjust project/programme implementation?”

Clearly, modifying and improving implementation is sharply different from assessing whether the activity is doing the right things. Attention to relevance becomes attention to effectiveness and efficiency. A review of evaluation reports suggests that this slippage is common. The consequence is that evaluators report that activities are relevant because they are found to have been done effectively or efficiently.

Relevant: Almost Always

Are the development activities Sweden supports relevant? Almost always, it seems.

Sida's annual evaluation reports are consistent on the evaluation of relevance in Swedish development assistance.

Sida's 2019 Evaluation annual report:

All evaluations that assessed an intervention's relevance found them to be relevant, with many finding the interventions to be highly relevant (p. 6).

Sida's 2018 Evaluation annual report:

Relevance and effectiveness are assessed in all evaluations. For a majority of the evaluated programmes, relevance was assessed as high (p. 11).

Sida's 2017 Evaluation annual report:

For a majority of the evaluated programmes, relevance was assessed as high (p. 14).

Sida's 2016 Evaluation annual report – examples:

The summative and formative evaluation of . . . finds that it is relatively relevant to the national education strategy (p. 14).

The evaluation concluded that the design of . . . has been relevant in relation to local needs, the national and regional environmental and climate change policies, and to the cooperation strategy of Sweden in . . . (p. 15).

The evaluation concluded that the projects have been largely relevant to [the country's] priorities and beneficiary's needs (p. 17).

Sida's experience is not unique in this regard. A German meta-review of evaluations of development assistance in Afghanistan found:

None of the 148 evaluation reports judged that a project was not relevant (Zürcher, p. 16).

As the reviewer notes, since there are many needs in conflict-ridden Afghanistan, every supported activity may be deemed relevant. Still, here too, if all activities are relevant, then relevance has limited utility as an evaluation criterion.

These, of course, are summaries of more detailed documents. An informal review of complete Sida decentralized evaluations over several years, and selected samples from other agencies, finds a few that report differentiated and nuanced attention to relevance. In those, the evaluator has explained the basis for the assessment of relevance and provided substantive evidence. In some of those the evaluator has reported that as implemented, an activity was more relevant to one referent (say, an objective stated at the outset) and less relevant to another (for example, an expectation articulated by local officials). Some evaluations provide systematic evidence that Swedish assistance has been attentive to needs,

policies, and priorities in the partner country at national and local levels (for example, Königson, Nilsson, Katende 2021). Those evaluations, however, are the exceptions. Most evaluations report on relevance in a more cursory manner, apparently seeing little need for explanations or evidence. The evaluation finds the activity relevant – in my extended but informal survey, never not relevant – and then moves on to what the evaluator takes to be more challenging issues. Relevance is assessed as dichotomous, not along a continuum, and there is no attempt to explore alternative referents for relevance.

How to understand the findings that all Swedish development cooperation is relevant or highly relevant? Read uncritically, those reports should lead to the conclusion that on that criterion Swedish support programmes are outstanding, indeed a model for others. Perhaps so, but even then while the assessment of relevance confirms that Sida's assistance is “doing the right thing,” its undifferentiated report cannot contribute to learning among development partners. As I have noted, most are superficial in their assessment. It is their confirmation of relevance that requires attention.

Evaluations of development assistance are generally complex and detailed. Where the assessment on a particular criterion is nearly always positive, especially when it is presented with little or no evidence, it adds little to the evaluation. Does the criterion or its assessment require major revision? Or should it simply be dropped entirely or reserved for particular evaluations where it can be expected to play a more prominent role?

Relevance as a Gatekeeper

Where relevance is understood as doing the right things, the relevance criterion becomes a gatekeeper at the entry portal to all the other evaluation criteria. From the funding agency's perspective, there is no point in supporting an activity that is not relevant, that is that is not doing the right things. The assisted activity thus begins with an official affirmation of relevance. Efficiency, effectiveness, and impact can all be expected, but cannot yet be confirmed. For evaluators, too, if an activity is not relevant, then why spend time examining whether the wrong thing has been done efficiently? The initial presumption of relevance is clear to evaluators.

Understood as a gatekeeper, a first check before proceeding with the substantive and demanding work of the evaluation, the relevance criterion encourages formulaic responses. As well, for evaluators, reporting to the funding agency that its assistance has gone wrong is a big hill to climb.

Next, it is important to focus attention on the implementation of the relevance criterion.

Relevance: Impossible to Assess

With the funding decision, relevance is presumed. How, then, are evaluators to proceed? How can relevance be assessed and confirmed?

An example is helpful. Especially in the immediate post-colonial era but occasionally in the present, the observation that most people in Africa are rural agriculturalists leads to the assertion that relevant education must focus on the tools and skills of farming. Unemployment is attributed to miseducation, that is, to studying history and language rather than soil chemistry and accounting. That education is not relevant. From that perspective, schools that teach languages to introduce young people to other cultures or assign books intended to expose learners to new ideas and different ways of thinking or insist that students use microscopes to understand and master systematic observation and comparison are wasting time with irrelevant programmes. Yet, if so, how will Africa ever escape its dependence on the ideas and technologies of others? How will Africa move beyond exploiting non-renewable resources to creating and developing new resources? If no Africans experiment with subnuclear particles, write new computer programs, use large data and remote sensing to improve agriculture productivity, or devise new approaches to dysentery, malaria, and AIDS, how can Africans assume responsibility for their own direction? Schools try to solve that problem by tracking – most students learn about farming, while a few have access to deeper challenges – but that is self-limiting, since it excludes the majority of learners from the generative process of creating new knowledge.

What, then, is relevant? If education is to expand rather than limit horizons, if education is to be developmental, both for individuals and for society, designing a relevant education must address multiple objectives

that intersect, that lead in different directions, and that are sometimes in tension. What is the appropriate referent for the relevance criterion?

Sida's Relevance Criterion

To bring the implications of that example to the present, consider Sida's specification of relevance (essentially identical to the OECD specification):

“The extent to which the intervention objectives and design respond to beneficiaries' global, country, and partner/institution needs, policies, and priorities, and have continued to do so if/when circumstances have changed.”

Read as written, the relevance specification expects evaluators to assess the “extent to which” an activity's (A) objectives and (B) design “respond to” the (1) needs, (2) policies, and (3) priorities of (a) beneficiaries at the global level, (b) beneficiaries at the country level, presumably (c) beneficiaries at the local level, (d) partners (Sida, perhaps others involved in the same activity), and (e) other institutions. That requires well-grounded entries in all of the cells of two versions of at least a 5x3 table:

Objectives

	Beneficiaries' Global	Beneficiaries' Country (National)	Beneficiaries' Local	Partner(s)	Civil Society Organizations
Needs					
Policies					
Priorities					

Design

	Beneficiaries' Global	Beneficiaries' Country (National)	Beneficiaries' Local	Partner(s)	Civil Society Organizations
Needs					
Policies					
Priorities					

Comparison over time requires several dated iterations of these tables.

As well, while Sida's specification refers to an activity's objectives and design, evaluators will generally be more concerned with its implementation, effectively creating a third table.

Implementation (Process and Outcomes)

	Beneficiaries' Global	Beneficiaries' Country (National)	Beneficiaries' Local	Partner(s)	Civil Society Organizations
Needs					
Policies					
Priorities					

Reorganizing the tables to include the basic and supplementary evaluation criteria and Sida's priorities would expand the tables, adding many more cells to address, and making the assessment still more complex.

Clearly, developing entries for all the cells of multiple iterations of those enlarged tables would be beyond the reach of even the most extensive evaluation. In the absence of further guidance, relevance is too vast, too complex, and too conflicted to assess.

How to make an impossible task manageable?

Relevant by Definition

For evaluators, one strategy for making that task manageable is to simplify. The starting point is the funding agency's presumption of relevance.

At first glance, the similar assessments – the evaluated activity was relevant – over many different and different sorts of projects, across settings and countries is puzzling. A systematic review of a diverse set of evaluations found an implicit assumption and a common thread, both leading to the same evaluation conclusion.

The assumption: if Sida provided the funding, then Sida must regard the activity as relevant. Indeed, so too for other funding agencies. From the perspective of evaluators, that assumption is not unreasonable. Accordingly, their starting point is that they are assessing an activity

assumed to be relevant. If, notwithstanding OECD-DAC's and Sida's guidance about context specificity and alternative referents, evaluators understand their task to be to assess relevance dichotomously, unless they see compelling evidence that the funding agency has grossly misdirected its support, they will confirm relevance. That is, rather than developing a critical assessment of relevance, evaluators are attentive to clear indications of irrelevance. Observing none, their assessment is straightforward: relevant.

The common thread: even more powerful is a definitional approach to relevance. Evaluators take relevance to mean consistency with a formally stated policy or other official indication of goals, objectives, or priorities. For example, if the country has committed to education for all, then all education activities are, by definition, relevant. Even activities one step removed from direct education assistance, say, support to libraries, or to school feeding programmes, or to non-school vocational skills development, can be assessed as relevant by definition.

If Sida's priorities are included, then the definitional scope is even wider. Activities that can be reported to be contributing to gender equity, or democratic practice, or environmental sustainability can be assessed as relevant.

Evaluations of support to skills training programs shows how relevance by definition can be fundamentally misleading. Across Africa, especially during the immediate post-colonial period development assistance supported youth skills training, for example, carpentry, masonry, and tailoring. Officials and educators declared those programs relevant, because the high-priority objective was to reduce youth unemployment: skills training is relevant to reducing youth unemployment, and thereby, to reducing poverty. An observer could note that (a) the local community has no need for and cannot soon employ 50 carpenters, masons, or tailors, (b) the newly trained carpenters, masons, and tailors have no start-up capital to purchase tools and materials, and therefore no effective way to use their new training to earn an income, and (c) the newly trained carpenters, masons, and tailors cannot demonstrate on-the-job practical experience to prospective employers. The observer would then make the general point that the training programs are starting at the wrong end: the community's problem is a demand problem, not a supply problem. More

training programs cannot solve the lack of jobs. Therefore, those training programs are reasonably deemed irrelevant to the objective of increasing youth employment, because they cannot accomplish that objective. Better, then, the observer would note, to use the funds to support, say, a housing development program, aimed at helping people build houses, which would both increase the housing stock and create a demand for carpenters and masons. The point: a country's commitment to education does not make every education programme relevant. Note as well that the apparent benefit and the claim to relevance may function to obscure institutionalized but unremarked gender bias, in this example, carpentry and masonry for boys and tailoring the option available to girls. Put sharply, an activity that can be deemed relevant by definition might prove to be quite irrelevant to its stated objective, reducing youth unemployment, and to its assumed broader goal, reducing poverty.

An impossibly complex assignment that requires multiple differentiated (“extent to which”) assessments is managed by converting it into a single yes/no finding. Combined, the assumption that Sida or another funding agency would not fund projects it does not regard as relevant and the broad compass of relevant-by-definition make assessing relevance a quick task for evaluators before they move on to the challenges of assessing efficiency or impact.

Relevant by Evaluator's Choice

For evaluators, an alternative strategy for managing the complexities of OECD-DAC's and Sida's specification of the relevance criterion is to choose and apply one or another of the referents that fall within its ambit. Since many referents are plausible and defensible, none needs extensive discussion or justification.

OECD-DAC's and Sida's guidance suggest that evaluators can determine which cell in the very large table imagined above – each a potential point of reference for assessing relevance – they will address, though evaluators' terms of reference may constrain that discretion.

Where evaluators do choose among alternative referents for assessing relevance, generally they do not explain why the selected referent is appropriate or the most important. Nor is that referent selected in a

process that permits comparison across activities and evaluations and thus learning from experience. One evaluation confirms relevance by congruence with one national policy, while another evaluation chooses a different national policy as its referent, neither explaining why that policy is the appropriate referent. A third evaluation refers to a Sustainable Development Goal, or a Sida priority, or a more narrowly stated objective of the specific activity, or a larger problem (environmental degradation, global warming). In this evaluation chaos, even a claim of irrelevance can be countered by choosing a different referent.

While each referent seems reasonable in its own terms, the collective result is cacophony, not coordination. Where referents are chosen at convenience from a very large set of possible referents, there can be no productive cumulation to support learning from experience and improving practice.

Even more problematic, broad referents for relevance create rationales even for activities that seem sharply at odds with the priorities of Swedish development assistance. For example, reducing taxes on wealthy people can be assessed as relevant to reducing poverty on the expectation that wealthy people will invest some of the retained income in new production, thereby creating jobs, and thereby reducing poverty.

To assess even part of what is specified, evaluators need guidance, either on which dimensions of relevance are of most interest to Sida, or on how they should decide which dimensions to review and assess. Evaluation Terms of Reference rarely, if ever, provide that clarity.

Explaining the Selected Referent: Complicating Challenges

Most evaluators who pursue this strategy simply point to what they regard as an appropriate referent and then use it to confirm relevance. The few who seek to explain their selection must address several important complicating challenges.

One is the gap between formal documents and actual practice. An education ministry's policy may specify instruction in English, but

systematic observation finds that nearly all teachers regularly rely on the local language. Should the relevance assessment be based on policy-made-in-practice (what teachers actually do and are expected to do) or by the policy-on-paper (which is widely ignored)? Can relevance be confirmed by congruence with a stated policy that is widely ignored and has little practical consequence?

Even more problematic is a second challenge. As written, the relevance criterion assumes a general harmony among needs, policies, and priorities and then across the specified points of reference. Yet, conflicts among them abound.

Consider support to teacher education. A country may *need* many more teachers quickly, to be achieved by reducing the length of teacher education programmes or favouring in-service over pre-service teacher education. At the same time, its *policies* may specify longer courses of study to achieve teacher certification and the accompanying salary. Its *priorities* may favour redirecting resources to measles vaccination and malaria eradication. Quite simply, within the country, needs, policies, and priorities are regularly in tension. What, then, is relevant? An activity that supports, say, stated policies and ignores expressed needs and articulated priorities? Or is an activity more relevant if it favours needs over policies?

Continue the example and add Sweden's needs, policies, and priorities. If an activity reflects an instrumental approach to achieving gender equality rather than a rights approach, is it to be assessed as not relevant? If an activity reduces gender inequality but is not transparent and has weak accountability, is it to be assessed as not, or less, relevant?

More generally, local demands (needs) are regularly in conflict with national policies and priorities. On important issues, a development partner's policies and priorities may differ from Sida's policies and priorities, sometimes sharply. Both may clash with the policies and priorities of a teachers union, or those of an association of homeless people and shack dwellers, or those of Oxfam, Save the Children, or another international non-governmental organization. Where they clash, which – whose – needs are to be the reference point for relevance?

Where needs, policies, and priorities are in tension, where a development partner and Sweden have divergent policies and different priorities, and where a particular activity and its institutions are buffeted by those tensions, what is relevant?

A third assessment challenge is determining whose perspective is to have greater weight in the evaluation process. In addition to determining whose needs are to have priority, evaluators must also determine whose assessment of those needs is to be regarded as authoritative. Where different participants in the assistance activity differ on needs (and policies and priorities), that determination is both difficult and consequential. I shall return to this issue.

What we see here is an almost unbounded set of possible strategies for assessing relevance. Evaluators can report that an activity is relevant because it is consistent with an important document, or consistent with a stated goal, an SDG, or a broad objective like gender equity. Or evaluators can find an activity relevant because officials said that it was relevant, or, less often, because activity participants said it was relevant. Or evaluators can point to outcome evidence, say, increased female enrolment or a decline in infant mortality, as confirmation of relevance. In the absence of explicit guidance on how relevance is to be assessed, evaluators not only regularly confirm relevance, but they do so in idiosyncratic ways that preclude comparison across activities and over time.

Note that while OECD-DAC's evaluation guidance anticipates and points to some of these challenges, often that guidance is not reflected in terms of reference for evaluations and not visible in evaluators' reports.

In sum, Sida's (and OECD's) specification of relevance is a complex terrain with many actors. The referents for relevance and therefore relevance are regularly and often sharply contested. That creates a foreboding arena that evaluators will generally seek to avoid, preferring the clarity of relevant-by-definition to the major effort required to construct an understanding of relevance that is sensitive to local conditions and responsive to the international environment and then to measure and document it.

Relevance is not a Fixed Target

What have we learned thus far? For Sida and some but not all development assistance agencies, relevance is an important evaluation criterion. The assistance review and approval process result in a presumption of relevance. That differentiates relevance from other evaluation criteria, where a positive assessment can be expected but not presumed. As well, evaluators have a compelling interest in finding relevant the activity they are evaluating. If the activity is not relevant, then why waste time evaluating its efficiency or sustainability? A review of Sida evaluations over recent years shows that nearly all evaluated activities were found to be relevant. So too for evaluations of other countries' development assistance. Accordingly, funding agencies can confirm that what they are doing is relevant, that they are doing the right things. Still, beyond that legitimizing confirmation, an evaluation criterion with no variation adds little to the evaluation.

Across different evaluators, different activities, and different countries, we have found, with some exceptions, three major approaches to assessing and then confirming relevance: relevance by assumption, relevance by definition, and relevance by selecting referents from a complex and chaotic cauldron.

To understand better why the relevance criterion is so often treated so superficially, it is important to consider the challenges of applying it analytically, synthetically, and critically. Relevance is not a fixed target.

Relevance is necessarily a relational criterion

Relevant to what? Critical here is the referent. As we have seen, in the absence of explicit guidance, evaluators can select whichever relevance referent they deem appropriate. A review of evaluations indicates that only very rarely do evaluators explain their choice of referent or show how that referent enriches the evaluation.

Explaining the choice of referent, however, could strengthen the evaluation. Why should relevance be assessed in terms of a formally stated national policy rather than, say, in terms of locally expressed needs? Or

why does a local organization regard the national policy or its implementation as irrelevant to local needs? Or alternatively, why should local priorities – as in: we need a school more than cattle dips or training in managing cooperatives – be ignored in assessing an activity’s relevance?

In practice, «relevant to what?» always becomes «relevant to whom?». Whose needs should have the highest priority? Why theirs? Whose policies should determine relevance? Why theirs? Whose priorities should be used to distinguish between what is and is not relevant? Why theirs?

Combining attention to relevant to what and relevant to whom creates space for converting the relevance assessment from a dichotomy to a continuum – from relevant/not relevant to more/less relevant. That in turn enables and encourages evaluators to explain why an activity may be assessed as effective but at the same time less relevant to local priorities than an alternative activity for which the community seeks funding.

Explicitly addressing relevant to what and relevant to whom could and should strengthen the assessment of the evaluated activity and in doing so could add substantive content, rather than mechanical affirmation of the relevance criterion.

Relevance is always time constrained

Needs, policies, and priorities change. Sida’s guidance is clear on that. Assessing relevance must reflect that.

For example, rapid recruitment and limited pre-service education may be relevant to a crash program to expand school access. As universal access is achieved, that approach may become irrelevant as the need for new teachers declines, priority to shifts to education quality, and policy specifies longer pre-service teacher education. Similarly, an epidemic or conflict may rearrange priorities that are again changed when the crisis has passed. Accordingly, a useful assessment of relevance must be time-stamped, noting the situation and circumstances of the determination, and examining whether or not an activity was modified as circumstances changed.

Assessing relevance always requires attention to who makes the assessment

Recall that where evaluators report their assessment of relevance hardly ever do they explain how they have made that determination. Effectively, they are reporting their professional judgment but not explaining the foundation for that judgment. Most important, evaluators are not considering how the assessment might differ, depending on who makes the assessment.

If the goal is to ask whether or not Sida is “doing the right things,” whose determination of what are the right things should guide the assessment? Surely it is likely that different sections of the Swedish government, different sections and levels of the partner country government, organizations engaged to implement a supported activity, and those directly involved in the activity will differ on what are the right things and thus on how resources should be allocated and managed.

Although the political context within which an activity was developed and implemented may not be a major focus for evaluators, ignoring that context makes it difficult to explore trade-offs between different needs, policies, and priorities and negotiated understandings between Sida and its development partners.

For water resource management, for example, some farmers might find the prospect of increased irrigation directly relevant to their needs. Other farmers, who are to be displaced by the construction of a dam or diversion of a river, might find the same activity inconsistent with their needs, not “the right thing.” Education activists might find the water resource management activity beneficial but of low priority and largely irrelevant to what they regard as the most pressing societal needs, and thus far from “the right thing.”

An evaluator who ignores those differing assessments misses the opportunity to evaluate the activity in its conflicted context and to give voice to different segments of the community. That missed opportunity may make it more difficult to understand why, for example, a community has not asserted ownership over an apparently successful activity and has not continued it beyond the end of the development assistance.

Failure to address «whose assessment of relevance is this?» limits the utility of the relevance criterion, weakens the overall evaluation, and constrains its ability to contribute to the formulation and implementation of development policy.

Relevance Assessments: Observations

We have seen that relevance is a unique evaluation criterion. Recall that while there can be an expectation of effectiveness or efficiency or sustainability, development assistance begins with a presumption of relevance. Sida and other funding agencies will not fund activities they deem irrelevant. Evaluators then begin their work with that presumption, creating strong pressure to confirm it. As well, relevance plays a gatekeeper role for all the other evaluation criteria. Put sharply, if a particular activity is not relevant (doing the right thing), then why focus on the effectiveness or sustainability of doing the wrong thing? For most evaluators the relevance criterion is dichotomous: an activity is or is not relevant. If they cannot confirm relevance, the rest of the evaluation serves little purpose.

Still, all involved in the evaluation of development assistance prefer to retain the relevance criterion, notwithstanding the challenges in implementing it. OECD-DAC and Sida have developed guidance for the relevance criterion, but there seems to be a major gap between that work on guidance and the terms of reference developed for evaluators and then their evaluations. How, then, to move the implementation of the relevance criterion from the nearly universal and largely unuseful confirmation to an informative and productive evaluation component?

Several summary observations stand out in this review of relevance as a criterion for the evaluation of development assistance.

First, as the review and examples thus far have shown, as assessed in evaluations, the relevance criterion has limited utility. Most often, the assessment of relevance is close to marking a tick box, based on a superficial comparison of program objectives and formally stated national objectives or priorities. With that approach, combined with the presumption of relevance that flows from the original funding decision, it is not surprising to find that most programs are assessed to be relevant.

Second, there is frequent slippage between relevance and other evaluation criteria. Evaluators use the effectiveness or impact of an activity as evidence of relevance. That empties relevance of content, converting it into a derivative description, rather than an independent evaluation criterion.

Third, that someone said so, whether a senior official or a participant in a funded activity, is at best weak evidence of relevance. More often, that largely serves to legitimize the evaluator's judgment, that is, to add credibility to the tick box. To support their assessment of relevance evaluators must go beyond an interviewee's comments to present and review the evidence that supports the assessment.

Fourth, the effective assessment of relevance must be an active and critical process, attentive to the issues noted above, not simply the report of an evaluator's professional judgment. Consider the other evaluation criteria. There are major debates about the sort of evidence necessary to support a positive assessment of impact, but there is widespread agreement that claiming impact requires evidence. So too for efficiency, effectiveness, and sustainability. The evaluator's assessment, whether positive or negative, remains unconvincing in the absence of evidence. No less so for relevance. An assessment of relevance that does not reflect and document a systematic and critical examination of evidence cannot be persuasive.

Fifth, the assessment of relevance can be and perhaps often is contested. Relevance is a function of situation, circumstances, context, perspective, and timing. Especially important is the assessment, or perhaps divergent assessments, of relevance by a development activity's intended beneficiaries. To be useful, an evaluator's assessment of relevance must incorporate consideration and analysis of conflicting understandings of relevance and conflicting assessments of the activity being evaluated.

Sixth, if it is to be more than a formulaic affirmation, relevance is difficult to evaluate. Exploring relevant to and for whom requires an imaginative evaluation plan that is sensitively executed. To support local ownership and sustainability, the assessment of relevance must hear and listen to discordant voices. Far from the simplicity of a tick box, assessing relevance is as demanding as assessing efficiency or impact.

Seventh, relevance is difficult, and likely expensive, to measure. For evaluations where reporting depends on measurement, there must be explicit attention to measuring relevance. That requires conceptualizing relevance as a variable. Not only relevant to whom, for whom, and in whose eyes, but also more or less relevant, with appropriate scales and indicators. It may be that efforts to measure relevance require time and resources that do not yield corresponding significant additional information about the activity evaluated.

Eighth, for the reasons outlined, determining what is relevant requires not a simple statement of the obvious but an on-going engagement with values, expectations, and constraints. A well-grounded assessment of relevance requires collaboration and negotiations with intended beneficiaries to ensure that the assessment reflects their understandings.

Relevance: Remove, Revise, Reinvigorate

The starting point for this review was frustration with the relevance criterion in evaluations of development assistance. Across activities, countries, and evaluators, nearly all activities were found to be relevant. Thus far I have reviewed briefly the OECD-DAC origins of the relevance criterion, Sida's specification and additions, dimensions of the relevance criterion, and challenges for assessing it. The next step is to use the critical observations to address the future of the relevance criterion in Sida's decentralized and partner-led evaluations.

Remove

While the OECD-DAC evaluation framework is widely used, it is not universal. Several major organizations involved in development assistance do not highlight relevance as a major focus of all evaluations. Evaluators are expected to organize their assessment around the funded activity's major objectives. They will be attentive to effectiveness, efficiency, impact, and sustainability, but they may not explore whether or not the activity has strengthened democratic practice or reduced gender inequality. Most often they do not report whether or not an evaluated activity is relevant.

Since relevance is a troubled criterion that is difficult to capture, and since most evaluators report that whatever they are evaluating is relevant, the relevance criterion does not contribute to learning from experience or to improved results-based management. *One approach is to remove the relevance criterion from the list of high priority focuses for post-activity evaluations.*

Sida is clear that there should be flexibility in designing evaluations and that it is not necessary to include all the evaluation criteria. Evaluators can be asked to consider relevance among other concerns but not required to report on it. Effectively, that approach shifts initial attention to relevance to an earlier point in the development assistance cycle: during the consideration and formulation of activities to be funded. *Assessing relevance becomes a task of aid approval rather than aid evaluation.* Relevance can function as a funding filter – in practice, that is already the case – and perhaps as part of a pre-funding review of proposed activities. A positive funding decision is confirmation that Sida regards the activity as relevant.

Sida might then identify selected evaluations in which relevance is to be assessed. For those, Sida could elaborate on the relevance criterion, specifying exactly what is intended and how it is to be assessed.

Revise

Alternatively, Sida can retain the relevance criterion, revising its content and implementation to make it more informative and productive. As they review funding requests Sida and its partners do want to consider which activities are relevant. While Sida has already deemed an activity relevant in its decision to fund it, evaluators can offer an independent review of that determination, based on their analysis of a funded activity.

Over many years Sida has invested major effort and energy in collaborating to develop the OECD-DAC evaluation criteria. Independent modification of the relevance criterion will be institutionally challenging, though within the spirit of the flexibility that OECD-DAC guidance suggests. Sida has periodically innovated and led its partner development assistance agencies. Revising the relevance criterion can be a new opportunity for an imaginative initiative.

My review of the relevance criterion leads to several recommendations for revising its specification and implementation. Since OECD-DAC is currently revising its guidance to evaluators, I have assumed that Sida does not anticipate its own revision of its evaluation criteria in the near term. Accordingly, while I think there is a compelling case for revising the relevance criterion, I have focused here on elaboration and implementation.

The recommendations below build on OECD-DAC and Sida guidance to date on implementing the relevance criterion. Though the guidance itself is rich, the path from guidance to practice is fractured. Applying that guidance requires unpacking the relevance criterion and reassigning responsibility for assessing it, as well as clearer explanations and expectations for evaluators.

1. Revise the relevance criterion to address the problems outlined above

As I have said, I presume Sida does not wish to depart from the OECD-DAC evaluation criteria. I include this recommendation simply to note the compelling case for a major revision of the relevance criterion for evaluation, both content and application. A significant revision must address the presumption of relevance, the equation of relevance and doing the right things, the complexity of the criterion's specification, and the attention to objectives and design without explicit corresponding attention to implementation process and outcomes.

2. Locate the primary responsibility for assessing relevance in the aid approval process rather than in aid evaluation

Sida will not fund development activities that it regards as irrelevant or unimportant (not “doing the right things”). The relevance criterion refers not to implementation but to objectives and design, which can be assessed before the activity has begun. As well, as drafted and presented, the relevance criterion is very difficult, or perhaps impossible, to evaluate fully, with supporting evidence. As we have seen, evaluators respond with a formulaic affirmation that evaluated activities are relevant. An initial remedy is to relocate primary attention to relevance to the process of approving funding.

Doing so would distinguish more clearly between what an activity will address (which can be assessed pre-activity) and what an activity has accomplished (which can be assessed only post-activity). Sida can and does determine prior to approval that a proposed activity is consistent with a national policy and thus relevant. There is no need for a post-activity evaluator to say that again.

If evaluators find that the supported activity was implemented in ways that departed from that initial confirmation of relevance, they can of course address that. Recommendations for the post-activity evaluation are below.

3. Revise the interpretation and implementation of the relevance criterion to distinguish clearly between relevance (what is related and appropriate) and “doing the right things” (what should be done)

It is useful to return for a moment to OECD-DAC’s and Sida’s equation of relevance and “doing the right things.” As I have noted, in everyday language in English those are distinct ideas. For example, consider activities intended to improve science education at the primary level. Among the activities are updating science textbooks, providing laboratory equipment and instruction on its maintenance, and supporting science fairs and other opportunities for students to experiment and demonstrate their results. All might be assessed as relevant to improving science education. But are they doing the right thing? Perhaps not. On examination, some of those activities might embed the common assumption that girls are unable to do well in science. Relevant, but not right. A fuller examination might indicate that the primary obstacle to improved science education is teachers’ own limited science mastery and therefore conclude that teacher education is the most important activity. All of those other activities are then relevant but not right.

The common evaluation report that an activity is consistent with a national policy document neither examines relevance critically nor addresses doing the right things.

Thus, the common understandings of those two terms risks evaluation confusion. Is the evaluator to assess whether or not a supported activity is relevant to – consistent with – a particular objective or goal? Or is the evaluator to offer a broader assessment of whether or not Sida is doing what it should be doing?

In the absence of clear directions, evaluators must wrestle with what, exactly, Sida seeks. Different evaluators may frame their assessments differently, making comparisons across activities and over time difficult or impossible. To make the relevance criterion useful, Sida must indicate to evaluators whether it seeks an assessment of relatedness or correctness or both. If relatedness, relevant in relation to which objectives or goals, or alternatively, what process is to be used to determine which objectives or goals are to be the relevance referent? If correctness, in whose terms?

4. Adopt an active posture in differentiating among types of evaluation and reserve the evaluation of relevance to those expected to contribute to the general evaluation of Sida's development assistance

Sida's Evaluation Handbook recognizes that evaluations of development activities serve multiple purposes and several audiences and that evaluations should be utilization-focused. Most common, however, is a broader evaluation intended to serve several purposes simultaneously. This recommendation suggests that Sida implement explicitly its guidance on utilization-focused evaluations.

An evaluation that functions to support continued funding or to show impact may serve those purposes but be of little use to those involved in a supported activity, for example early childhood teachers working on curriculum development. Similarly, an evaluation whose primary purpose is to enable early childhood teachers to become more skilled at curriculum development may not serve well to assess outcomes or to report to Sweden's parliament. Sida can and should differentiate more clearly among the evaluations it seeks. That would permit greater clarity from the outset on which evaluation criteria should be addressed and which approaches and methodologies are appropriate and likely to be informative to the intended audience.

As well, doing so enables Sida to designate a few evaluations, expected to contribute to a broad assessment of Swedish development assistance, as appropriate for the evaluation of relevance. Those evaluations may be central or decentralized, larger or smaller scale, and focused on a broad set or a single activity. Where an evaluation is expected to assess relevance, it must include significant participation by those responsible for the activity being evaluated and must consider alternative assessments of relevance.

5. Modify Terms of Reference for evaluations expected to explore relevance to specify which dimensions of relevance are to be evaluated, with explicit attention to relevance for whom and to whose perspective is to guide the relevance specification

Since relevance is a complex criterion, Terms of Reference for evaluations intended to assess relevance must provide fuller guidance on Sida's expectations. The message to evaluators must be clear: assessing relevance is demanding and like assessing effectiveness or efficiency, requires imagination, insight, and evidence. Sida must indicate which dimensions of relevance are to be given priority. The assessment of relevance must include attention to relevance for whom and a critical review of alternative perspectives on relevance. It must as well must examine relevance as a continuum rather than a dichotomy. Accomplishing that more demanding assessment of relevance will require a consultative evaluation design and strong participation by those involved in the funded activity.

Other evaluation criteria, especially those to which Sida assigns high priority, can be addressed as appropriate.

6. Recognize that relevance can be and often is a contested assessment and explicitly expect evaluators to review alternative assessments of relevance

Most often, when evaluators assess relevance, they are reporting their own professional judgment, generally with little supporting evidence. The challenge is to strengthen that assessment by recognizing that there are likely divergent views, explicitly exploring alternative perspectives, and providing supporting evidence for the final assessment.

7. For evaluations primarily intended to enrich and extend the funded activity, the Evaluation Terms of Reference should expect participatory evaluation, from conception and design of the evaluation through its implementation and report. Relevance will not be a primary focus and may not be addressed at all

Within Sida's guidance on utilization-focused evaluation, some evaluations may be designed primarily to strengthen the supported activity, rather than to report to Sida on efficiency or impact. That should be clear in the evaluation's Terms of Reference. That clarity should include guidance on which among the six evaluation criteria must receive priority attention.

Although Sida's evaluation guidance refers to participatory evaluations, it seems they are rarely undertaken. Limiting the assessment of relevance to selected evaluations permits other evaluations broader latitude in approach and methodology. That enables Sida and its evaluators to develop and implement innovative approaches that permit both reporting what has happened and learning from experience. The advantages of participatory evaluations can prove far more consequential than their limitations.

8. Wherever it is specified as an evaluation criterion, relevance should be assessed as a continuum (highly relevant to not relevant at all) rather than a dichotomy (relevant/not relevant)

As we have seen, most evaluations report that the activities evaluated are relevant, implicitly understanding the possible assessments to be relevant or not relevant. To make relevance a more useful evaluation criterion, Sida and its evaluators should consider relevance as a continuum and expect evaluators to define the continuum, locate the evaluated activity along that continuum, and explain their assessment.

Evaluators may examine several dimensions of relevance. For example, was the activity, as implemented, consistent with a consequential national priority? Did the activity, as implemented, contribute demonstrably to reducing gender inequality? Did the design of the activity, as implemented, include components that demonstrably made it sustainable beyond the completion of the external funding? Each dimension should have its own continuum and supporting evidence.

9. Partner country evaluation institutions should have primary responsibility for post-activity evaluation of relevance

Sida is clear that its development partners should have major responsibility for evaluating supported activities. Evaluations "shall rely on [Sida's] partners' monitoring and evaluation systems to the greatest extent possible . . . responsibility for evaluation of Sida-funded programmes rests primarily with the development partner." (Sida's Evaluation Handbook, 2020: 9).

It is appropriate and productive, therefore, that just as pre-activity assessment of relevance should be relocated to the aid approval process, partner country institutions should have primary responsibility for post-

activity assessment of relevance. The goal is to assure that the primary referents for relevance are the needs, policies, priorities, and practices of the partner country, from national to local levels. A related goal is to encourage deep collaboration between assistance provider and assistance recipients on issues of high priority common interest.

The overall evaluation plan will need to incorporate this new element. As appropriate, Sida will need to assist partner country evaluation institutions in addressing this task. Transitioning to this arrangement will require experimentation and a commitment to recognizing and addressing disagreements and divergent understandings and priorities.

Where the partner country prefers, or where the partner country's evaluation institutions have limited capacity, Sida may commission an evaluation in the usual way. As it does so, it must be clear that its evaluators are expected to collaborate with partner country institutions and individuals to assure that the post-activity assessment of relevance is firmly grounded within the partner country setting.

As appropriate, within this understanding Sida may commission an evaluation directly focused on relevance.

Reinvigorate

A major challenge in making the relevance criterion useful is to transform it from a quickly addressed tick box to an informed, nuanced, and well-grounded assessment. Doing so requires not only differentiated evaluations and explicit guidance, but also modifying what apparently has become an evaluation cultural style: relevance as a tick box that does not require much attention. Changing the practice requires modifying expectations that have a long history and that are ingrained in approaches and understandings shared among evaluators. Changing the practice requires changing the cultural style.

It may be appropriate to design a few evaluations for which relevance is the primary or sole focus. Evaluators can be encouraged to be imaginative in that design, remaining attentive to the audiences for whom relevance is a high priority evaluation criterion.

Sida's out-sourcing evaluation strategy and its reliance on evaluation firms whose existence depends on future evaluation contracts creates a structural disincentive to report to Sida that it is not doing the right things. Sida of course welcomes the good news of a positive evaluation of relevance, but it must create space for news that is not so good, that is, a negative evaluation of relevance. Reinvigorating the relevance criterion will require communicating, both in words and in practice, that Sida seeks an independent and well-grounded assessment of relevance, which may include the report that all or parts of a funded activity are marginal to or at odds with Sida's or its partners' needs, policies, and priorities.



Currently, relevance assessments in Swedish development cooperation confirm relevance but yield little information that can contribute to improved management or implementation of development assistance. Most often, they are formulaic and definitional. Even where evaluators provide some details, their assessments are so idiosyncratic that they do not support learning from the experience and cannot contribute to comparison across funded activities.

Assessing relevance remains important. Unpacking the relevance criterion makes that possible. Reassigning responsibility for the assessment can both facilitate results-based management and deepen and strengthen partners' participation in the evaluation process.

References

- Cabria, Nicola, (2014). Does China Plan and Evaluate Aid Projects like Traditional Donors? Stellenbosch, South Africa: Stellenbosch University, Centre for Chinese Studies.
- Chianca, Thomaz. (2008). “The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement.” *Journal of MultiDisciplinary Evaluation* 5, 9: 41–51.
- Christoplos, Ian, Anna Liljelund Hedqvist and Jessica Rothman. (2013). *Swedish Development Cooperation in Transition? Lessons and Reflections from 71 Sida Decentralised Evaluations (April 2011–April 2013)*. Swedish International Development Cooperation Agency: Department for Evaluation. July 2013. *Sida Studies in Evaluation* 2013:1.
- Christoplos, Ian, Anna Liljelund Hedqvist and Jessica Rothman. (2014). *Lessons and Reflections from 84 Sida Decentralised Evaluations 2013 – a Synthesis Review*. Swedish International Development Cooperation Agency: Department for Evaluation. October 2014. *Sida Studies in Evaluation* 2014:1.
- Clements, Paul. (2020). “Improving Learning and Accountability in Foreign Aid,” *World Development* 125.
- Danielson, Anders, and Lennart Wohlgemuth. (2003). *Swedish Development Cooperation in Perspective*. Lund: Lund University, Department of Economics.
- Edgren, Gus. (2003). *Donorship, Ownership and Partnership: Issues Arising From Four Sida Studies of Donor-Recipient Relations*. Stockholm: Swedish International Development Cooperation Agency, *Sida Studies in Evaluation*, 03/03.
- Eduards, Krister. (2006). *Review of SIDA's research cooperation: Synthesis Report*. Swedish International Development Cooperation Agency: Secretariat for valuation and Internal Audit. November 2006. *Sida EVALUATION* 06/57.
- Forss, Kim. (2019). “Getting Value for Money? A Critical Analysis of the Costs and Benefits of Evaluation,” in Jan-Eric Furubo and Nicoletta Stame, editors, *The Evaluation Enterprise: A Critical View*. London: Routledge.
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. (2016). *Impact Evaluation in Practice*. Washington: World Bank, Second Edition.

- Glennie, Jonathan, and Andy Sumner. (2014). *The \$138.5 Billion Question: When Does Foreign Aid Work (and When Doesn't It)?* Washington: Center for Global Development.
- Königson, Åsa, Marie Nilsson, and Stephen Katende. (2021). *Evaluation of TGNP Mtandao (Formerly Tanzania Gender Networking Programme) Extended Strategic Plan 2016–2019 in Tanzania. Final Report.* Stockholm: Sida Decentralised Evaluation 2021.5/
- Organisation for Economic Cooperation and Development, Development Assistance Committee, Network on Development Evaluation. (2010). *OECD DAC Quality Standards for Development Evaluation.* Paris: OECD
- Organisation for Economic Cooperation and Development, Development Assistance Committee, Network on Development Evaluation. (2019). *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use.* Paris: OECD.
- Patton, Michael Quinn. (2020). *Blue Marble Evaluation: Premises and Principles.* New York: Guilford Press.
- Picciotto, Robert. (2020). "From Disenchantment to Renewal," *Evaluation* 27, 1:49–60.
- Samoff, Joel, Jane Leer, and Michelle Reddy. (2016). *Capturing Complexity and Context: Evaluating Aid to Education.* Stockholm: Expert Group for Aid Studies, 2016). [<http://eba.se/wp-content/uploads/2016/05/Rapport-201603-Capturing-complexity-and-context.pdf>]
- Swedish International Development Cooperation Agency. (2017). *Evaluations at Sida. Annual Report 2016.* Stockholm: Sida.
- Swedish International Development Cooperation Agency. (2018). *Evaluations at Sida. Annual Report 2017.* Stockholm: Sida.
- Swedish International Development Cooperation Agency. (2019). *Evaluations at Sida. Annual Report 2018.* Stockholm: Sida.
- Swedish International Development Cooperation Agency. (2020). *Evaluations at Sida. Annual Report 2019.* Stockholm: Sida.
- Swedish International Development Cooperation Agency. (2020). *Sida's Evaluation Handbook: Guidelines and Manual for Conducting Evaluations at Sida.* Stockholm: Sida.

United Nations Evaluation Group. (2016). Norms and Standards for Evaluation. (New York: UNEG).

United States Agency for International Development. (2016). Evaluation: Learning from Experience. USAID Evaluation Policy. Washington: USAID, Updated.

Vaessen, Jos, Sebastian Lemire, and Barbara Befani. (2020). Evaluation of International Development Interventions: An Overview of Approaches and Methods. Washington: World Bank, Independent Evaluation Group.

Zürcher, Christoph. (2020). Meta-Review of Evaluations of Development Assistance to Afghanistan, 2008–2018. Chapeau Paper. Federal Ministry for Economic Cooperation and Development (BMZ), Germany.