

Notes for *Basics of Evaluation*

AESIS 2023: Institutional Structures for Societal Impact of Science

May 23, 2023

50 Minutes

John Gargani

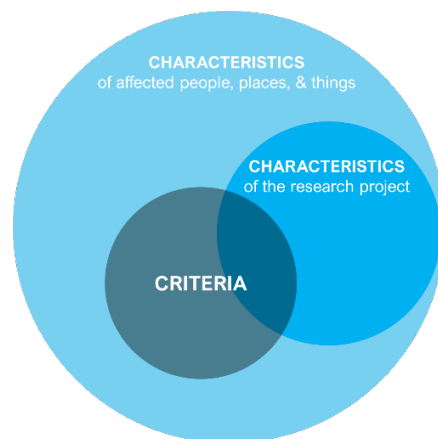
john@gcoinc.com

We must make choices about how to conduct an evaluation, and we must justify them to others.

The Definition of Evaluation

To evaluate is to **determine** the **value** of **something**.

The “something” may be anything we wish. In our case, it is likely science, research, researchers, research institutions, research proposals, etc. Whatever we choose to evaluate, we must be able to understand and describe it. However, it will have many characteristics. Some characteristics directly describe what we are evaluating, such as the cost of a research project. Other characteristics describe affected people, places, and things. These characteristics include the impact of science and research on society. We can only use some characteristics in an evaluation, and the characteristics we choose are called **criteria**. The results of an evaluation often depend on which criteria are chosen.



So, we must answer—and justify our answers to—the following questions:

What is being evaluated?

Which criteria matter?

Who decides?

Who decides who decides?

Value is an important topic to evaluators. There are different kinds of value—merit, worth, significance, meaning, etc. They may shed light on scientific value, economic value, cultural value, etc. Not only must we estimate how much of each type of value that science and research may have produced, we need **standards** to interpret the estimates. How much of what type of value is adequate? Excellent? Insufficient? Whose perspective on value matters? Some groups may capture some or all the value that science and research create. Others may lose value. Every group, and every person within a group, may value what has been accomplished differently, favoring different types of value and applying different standards of success. So, we must answer—and justify our answers to—the following questions:

What kind of value?

Value for whom as judged by whom?

What standards of value (what is enough, excellent, etc.)?

Who decides?

Who decides who decides?

Evaluators determine value by reaching **evaluative conclusions**. This should be done in a way that is systematic (it may be explained to others), logical (conclusions follow from evidence and reasonable assumption), and pragmatic (an explicit tool, method, or strategy is used to implement the logic). So, we must answer—and justify our answers to—the following questions:

What systematic logic?

Which practical approach?

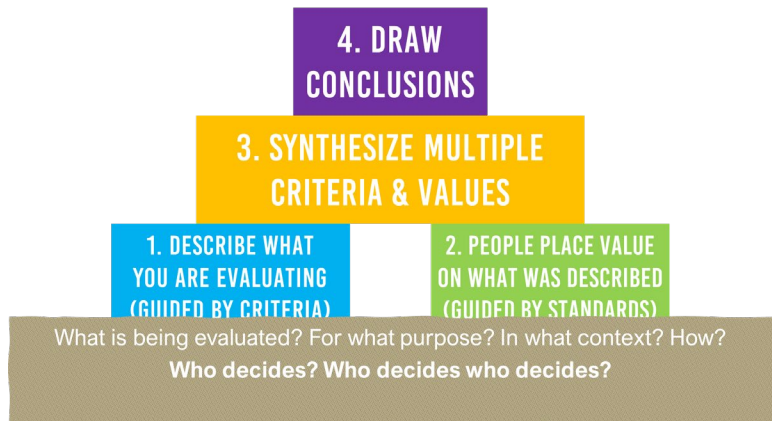
Who decides?

Who decides who decides?

The General Logic of Evaluation

The general logic of evaluation is one systematic logic (there are others). “Preliminaries” ground the evaluation. They establish what is being evaluated, why, in what context, and how. They also acknowledge who decides how the evaluation will be conducted and who has power. With this in place,

- (1) evaluators describe what is being evaluated as guided by the criteria,
- (2) people (those who are affected or have a legitimate interest) place value on what was described as guided by the standards, and then
- (3) evaluators synthesize multiple criteria, standards, perspectives, etc. in such a way that
- (4) they can reach a valid evaluative conclusion.



The distinction between description (box 1) and value (box 2) is fundamental to evaluation. What most people mean by “measuring impact” is description. For example, if we measure the impact of AI and find it has caused some jobs to disappear, we have described its impact. This is a “fact” of the world (that we may or may not have described well). However, people may differ about whether the impact is good or bad, and they may place more or less value on it. To evaluate impact, we must describe and value it.

Synthesis poses a longstanding problem. No one has developed an approach that works well all the time, in every context, and for every purpose. Often, it makes sense to report success on each criterion separately *and* aggregate them in some way.

Rubrics

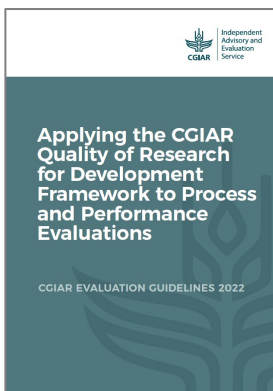
Rubrics are one tool among many for implementing the general logic of evaluation. They are often organized as a matrix with a column for criteria and another for standards. The more criteria under consideration, the more rows in the matrix. A separate rubric may be constructed for different constituencies, for example one for researchers and another for communities impacted by the research. This makes it possible to reach multiple (possibly contradictory) evaluation conclusions that represent multiple perspectives. Examples of rubrics used to evaluate science and research may be found in the first two references.

CRITERIA	STANDARDS

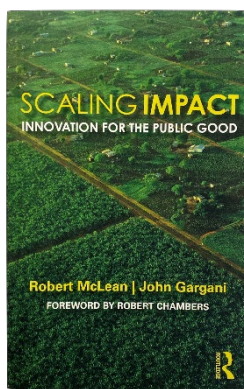
Free Resources (Except One)



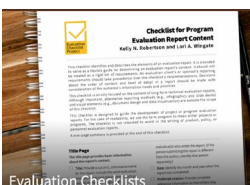
The International Development Research Centre (IDRC) in Canada funds research for development (scientific research undertaken to improve the lives of people and the environment in developing regions of the world). The organization developed a flexible approach to evaluating the research it funds called Research Quality Plus (RQ+). It is highly regarded, has been adopted and adapted by many research institutions, and was featured in the journal *Nature*. It proposes criteria for assessing research and the context in which it is conducted. Users are encouraged to modify, subtract from, and add to the criteria. A more detailed presentation may be found [here](#).



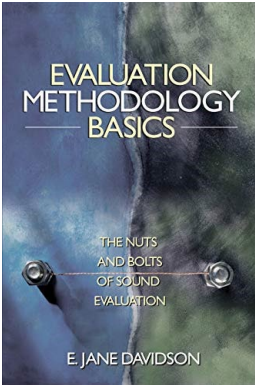
The Consortium of International Agricultural Research Centers (CGIAR) is the world's largest global agricultural innovation network. It recently developed flexible guidelines for evaluating research for development. The guidelines suggest combining CGIAR's criteria for evaluating "quality of science for development" with the OECD DAC criteria, which are widely used to evaluate international development programs.



Scaling Impact: Innovation for the Public Good is a book I wrote with Rob McLean, who works at IDRC on RQ+ and other evaluation projects. We conducted research to understand what innovators in the Global South consider successful scaling. We organized what we learned as four principles of scaling and included five case studies that explore how the principles work in practice. The book is part of a larger project to develop *scaling science*—a science of scaling that we hope will contribute to scaling the positive impacts of scientific research.



The Western Michigan University Checklist Project has produced several evaluation checklists. They describe good practice in areas such as evaluation design, reporting results, and evaluating evaluations (called meta-evaluation).



This is the only resource that is not free. Jane Davidson's book is a great starting place for anyone interested in evaluation. It describes how to thoughtfully implement the logic of evaluation in straightforward, nontechnical language. Available wherever you buy books.