

## OPEN ACCESS

Free to share and post this prepublication draft.

## CITE THE PUBLISHED ARTICLE

Gargani, J., & King, J. (2024). Principles and methods to advance value for money. *Evaluation*, 30(1), 50-68. DOI: 10.1177/13563890231221526

---

# Principles and methods to advance value for money

## John Gargani

Gargani + Company, Inc., USA  
<https://orcid.org/0009-0007-1409-6276>

## Julian King

Julian King & Associates, New Zealand

## Abstract

Value for money poses the question, “What is good resource use?” It is often answered with a narrow economic analysis that does not adequately address what diverse people value. We suggest new principles and methods that may help evaluators answer the question better. First, we define value for money, which sits at the intersection of evaluation and economics. Next, we make the case for a holistic assessment of value for money that evaluators can conduct with tools they already have, like rubrics. We introduce three principles that further align value for money with evaluation: value depends on the credibility of estimates; things do not have value, people place value on things; and people value the same things differently. Together, they suggest evaluators should arrive at multiple, possibly conflicting conclusions that represent diverse value perspectives. We demonstrate how this may be done using a value-for-money rubric to improve resource allocation for impact.

## Keywords

Value for Money, Valuing, Impact, Economic Evaluation, Rubrics

The concept of value for money is rooted in a commonsense idea of fairness—people should get what they pay for. Markets, when they work well, help ensure this fairness. Buyers are free to choose whether to make a purchase and have enough information to identify the best alternative available to them. Sellers are similarly informed, free to sell or not, and able to identify their best alternative. With many buyers and sellers competing under these conditions,

---

## Corresponding author:

John Gargani, Gargani + Company, Inc., 2625 Alcatraz Avenue, Number 508, Berkeley, CA 94705, USA.  
Email: [john@gcoinc.com](mailto:john@gcoinc.com)

money only changes hands when both parties judge the value of what they obtain to be greater than the cost of obtaining it. This constitutes a fair exchange of value for money.

Embedded within this idea of fairness is the acknowledgment that people value the same things differently. What one consumer considers a good price–quality–value combination for a meal, pair of shoes, or car, another may not. Consequently, the question “Is this a fair exchange of value for money?” does not have a single answer but many that reflect diverse value judgments made by individuals and groups. This complicates value for money in traditional markets, but only to a small degree. Economists can describe diverse value perspectives mathematically with demand curves and utility functions, and ultimately markets decide whether enough people find sufficient value in a good or service to make it economically viable.

However, the concept of value for money becomes more problematic when governments, philanthropists, investors, entrepreneurs, and nongovernmental organizations “buy and sell impacts” for the benefit of others. For example, the Public Services (Social Value) Act 2012 in the United Kingdom requires commissioners of public services to consider how they can secure wider social, economic, and environmental benefits. A common strategy is to base purchasing decisions on a weighted combination of price and non-price criteria, including the lasting impact a purchase will have on individuals, communities, and the environment, referred to as its social value (UK Government, 2020: 2–3, 15). Around the world, roughly 250 social impact bonds have been contracted (Brookings, 2023), each obligating public officials to purchase social and environmental impacts at a price that is consistent with the values and mechanisms of private-sector investors (Fraser et al., 2018). And in the United States, there are recent calls to create impact markets in which impacts may be purchased at prices that are set by a central organization using verified performance data (Saul et al., 2023). In contexts such as these, there are (at least) three parties who must be satisfied by an exchange—buyer, seller, and stakeholder—and the third typically has little control (Gargani, 2019; Laplume et al., 2008). The price and quality of impacts are more difficult to ascertain than those of consumer products, financial instruments, and other familiar market goods, and the act of “buying” an impact (or more accurately, the chance of an impact) encompasses grants and investments in addition to straightforward purchases. These complications stretch the definition of markets and affect their ability to ensure fairness. This has led some to question how well the concept of value for money promotes the public good (Destremau and Wilson, 2017; McKeivitt, 2015).

We contend that value-for-money determinations should play a central role in the allocation of resources for the benefit of diverse communities, but they must change to do so effectively. To that end, we present principles and methods to help evaluators conduct value-for-money studies that meet evaluation standards, take multiple value perspectives into account, and, if warranted, reach multiple evaluative conclusions. We begin by defining value for money as good resource use (King, 2017, 2019). This frees evaluators from “narrow” interpretations that equate value for money with return on investment and benefit–cost ratios, which tend to rely on a single quantitative metric, and situates value for money as part of a “holistic” assessment of value, by which we mean evaluative judgments based on a more comprehensive set of criteria and standards that adequately represent the perspectives of different stakeholder groups, in particular the people directly affected. Holistic assessments of value sit squarely in the domain of evaluation; consequently, evaluators can conduct a value-for-money analysis with tools they already have in hand. One widely used tool is

rubrics, and we use them to illustrate how two economic measures that may be considered sufficient under a narrow interpretation, costs and social benefit–cost ratios, can be treated as criteria alongside others to judge good resource use in a more holistic way. Importantly, equity, social justice, inclusion, and similar “soft” objectives may be among the criteria even though they are considered blind spots of economic analysis (see, for example, Klonschinski, 2014).

An advantage of using rubrics is that they may be constructed and interpreted in a participatory, inclusive manner, ensuring they reflect diverse value perspectives (Martens, 2018). Nonetheless, they are typically used to reach a single evaluative conclusion based on a negotiated framework of criteria, standards, and evidence. A value-for-money analysis may be better served by multiple conclusions that reflect the diverse value perspectives of groups and individuals. To achieve this, we introduce three principles (Gargani, 2017, 2018; Gargani and King, 2022) to guide evaluators conducting holistic assessments of value.

We preface this with a summary of underlying concepts that may be unfamiliar to some evaluators, and then provide an example of a value-for-money rubric that puts the definition and principles into practice. As we introduce each principle, we apply it to the rubric to illustrate how it supports a more holistic assessment of value. In the end, we are able to reach multiple evaluative conclusions that reflect how groups and individuals judge value for money differently. We conclude by discussing the strengths and weaknesses of a holistic assessment in relation to long-standing problems related to resource allocation for impact.

## Summary of underlying concepts

Value for money poses a question about how well resources are used. The way evaluators seek an answer depends on how they understand a number of underlying concepts. For evaluators trained in economics, finance, and accounting, these may be familiar. Many evaluators are trained in other disciplines, however, so we provide a summary to scaffold subsequent sections.

### *Value is more than money*

In everyday life and professional practice, value encompasses merit, worth, significance, quality, importance, and other similar constructs. People are surprisingly good at navigating the shades of meaning packed into the word. We can talk about the value of a friendship and the value of an investment, and listeners immediately understand we mean different things. When conducting an evaluation, we often narrow our attention to a few meanings that satisfy our purpose and context. A value-for-money analysis, for example, may focus exclusively on financial, economic, or monetary interpretations of value, and they often do to such an extent that many people incorrectly believe that value for money and return on investment are synonymous. However, defining value in economic terms is a choice, not a requirement. Nonmonetary resources, such as social capital and expertise, may be used to make a difference in the world, and people may value that difference in nonmonetary ways. Evaluators should incorporate as many different conceptions of value as needed to represent the diversity of perspectives held by stakeholders. In a value-for-money analysis, the salient question is often, “How well were monetary and nonmonetary resources used to create a difference in the world that stakeholders value in monetary and nonmonetary ways?”

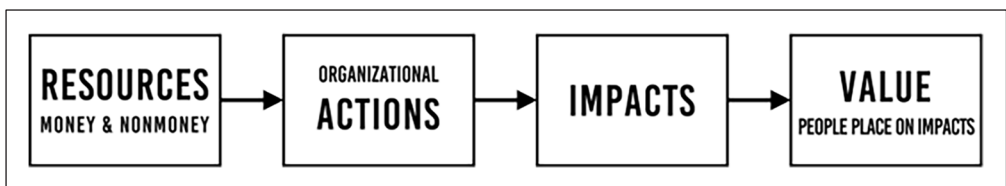
### *Impact plays a central role*

Evaluators routinely direct the question of value for money to programs, products, policies, investments, and grants. In this article, we limit ourselves to a feature that is common to all of them, impact. Impact is only one of many possible evaluative criteria, but it is typically among the most important because it would be difficult to justify that resources were used well in the absence of positive impacts. Evaluators define impact in different ways (Stern et al., 2012; see also Belcher and Palenberg, 2018; White, 2010). The definition we apply here subsumes the concepts of output, outcome, and impact used in logic models and theories of change. Impact is (1) all the differences found in (2) a specific group of people, places, and things at (3) a given point or span of time after they directly or indirectly experienced (4) two simultaneous and mutually exclusive actions undertaken by (5) one organization in (6) a given context. This is a counterfactual definition of the construct of impact. Note that it does not describe how the construct may be measured. In fact, impact by this definition may be measured using any method, qualitative or quantitative, experimental or non-experimental, that is available to evaluators. Part of an evaluator's job is to select an appropriate method given context, purpose, and constraints.

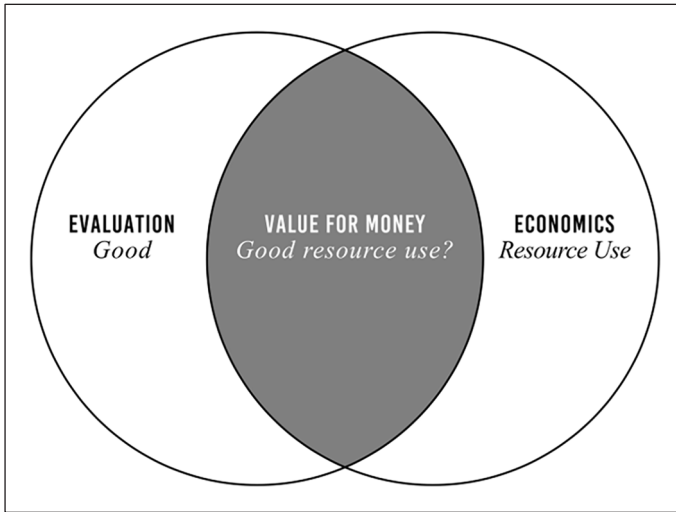
Measuring impacts as we define them is challenging. First, we cannot fully observe an impact. One organization cannot simultaneously implement two mutually exclusive actions. Nor can a person, place, or thing have mutually exclusive characteristics—a set of characteristics associated with one of the actions and another set associated with the other action (Holland, 1986). Second, we cannot attend to all the impacts a group of people, places, and things may experience. We can, however, describe some impacts to some level of approximation using a variety of research designs. The value stakeholders place on programs, policies, and other organizational actions depends in part on which impacts evaluators choose to describe and how well they describe them.

### *Value for money spans the entire causal chain*

We may situate the question of value for money in a causal chain that is familiar to evaluators—resources make it possible for an organization to take actions that produce impacts that people value (Figure 1). This schematic pathway hides many potential complexities, but it captures the net causal relationships that evaluators typically hypothesize ripple through a theory of change. Evaluators may investigate any combination of causal links in the chain to shed light on various aspects of efficiency and effectiveness. When answering the question of value for money, however, they must consider the entire span to ascertain the relationship between the value people place on impacts and the money and nonmoney resources required to create them.



**Figure 1.** Causal chain used to frame the question, “Is this good resource use?”



**Figure 2.** Value for money sits at the intersection of evaluation and economics and poses the question, “Good resource use?”

### *Value for money sits at the intersection of economics and evaluation*

Economics and evaluation are expansive disciplines. At their core, economics is the science of resource use, and evaluation is the science of judging what is good. Value for money sits at the intersection of the two, posing the question, “Is this good resource use?” (Figure 2). It is “an evaluative question about an economic problem” (King, 2017: 102). Importantly, we need not base our answer solely on economic criteria (e.g. economic efficiency or overall welfare), economic indicators (e.g. benefit–cost ratios, return-on-investment percentages, and net-present-value estimates) and economic standards of success (e.g. break-even points, hurdle rates, and positive net present values). Economic criteria and standards play an important role in value-for-money evaluations, but only a role because what makes resource use good for some stakeholders may not be economic in nature. It may, for example, include stewardship of resources, productive ways of working, and creating impacts with desirable qualitative characteristics like equity, sustainability, and variety (King, 2019; McLean and Gargani, 2019; Organisation for Economic Co-operation and Development (OECD), 2021) that purely economic methods may struggle to incorporate (Adler and Posner, 2006; Julnes, 2012; King, 2023).

Combining theory and practice from economics and evaluation answers the call from some economists to combine insights from economic analysis with “messier” aspects of democratic decision-making (Adler and Posner, 2006; Kahneman, 2012), in which economic criteria and evidence are “just one set of inputs that are amalgamated into the overall decision-making process” (Flyvbjerg and Bester, 2021: 356). To address this opportunity, an inter-disciplinary approach was developed, combining holistic criteria, standards, and evidence (King, 2019).

### *Value-for-money evaluations should use holistic criteria, standards, and evidence*

A set of criteria, standards, and evidence are “holistic” when they are sufficiently comprehensive to represent what diverse people value, but no more. This does not

suggest that a narrower economic analysis cannot sometimes suffice for a value-for-money analysis, or that economic analysis cannot widen its gaze. Multi-criteria decision analysis is one example of a more holistic economic approach (Dodgson et al., 2009). It was developed to guide decision makers who must balance competing objectives (Keeney and Raiffa, 1976) in decision-making environments that are too complex for cost-benefit analysis alone. Objectives inform the development of multiple criteria, only some of which can be achieved by any available option. Consequently, decisions pose trade-offs, and a single best solution may not exist. Some solutions may be better than others, however, and these are what multiple-criteria decision analysis identifies. This is done by assigning importance weights to criteria then calculating weighted sum scores (or another type of weighted score) for each decision option. Options with higher scores are better, and should an option have the highest score, it would be preferred above all others (King, 2019). In Europe, multi-criteria decision analysis has incorporated individual-level weights for small groups of “assessors” representing funders and other program partners, making it possible to arrive at separate conclusions for each (European Commission, 1999). In what follows, we suggest a more radical application of this approach by calling for the participation of the groups and individuals who are directly affected by resource use, and drawing evaluative conclusions for each.

Scriven (1991) argued that multi-criteria analysis, which he called numerical weight and sum, has limitations. Although “sometimes approximately correct, and nearly always clarifying” (p. 380), he argued it could lead to questionable evaluative conclusions. For example, many relatively unimportant criteria may dominate, leading to decisions that promote secondary or even tertiary objectives over the primary purpose of a policy or program. Scriven instead advised using a small number of “qualitative” ordinal categories, such as *high importance*, *medium importance*, and *low importance*, in which criteria may be grouped. Within categories, evaluators are free to aggregate evidence using unweighted sums, qualitative assessments of relative value, or other methods, but they should not aggregate across categories (Scriven and Davidson, 2000). He called this approach qualitative weight and sum (Scriven, 1994). Its justification rests on the tendency for criteria to be more similar within categories than across, making the synthesis of evidence within categories more tractable than synthesizing across all criteria. We apply the same logic to the aggregation of evidence within stakeholder groups, arguing that it may similarly ease the “synthesis problem” to some degree (see *Conclusion*).

### *The synthesis problem remains a problem*

Numerical and qualitative weight and sum are two strategies among many for addressing what is known alternatively as the aggregation problem or *synthesis problem* (Scriven, 1993). Whenever evaluations entail multiple criteria (multiple stakeholder groups, multiple sites, etc.), an evaluator must combine evidence in some way to understand the overall value of what is being evaluated. The more holistic these are, the more difficult the problem, but even narrow approaches to economic assessment like cost-benefit analysis struggle with synthesis (Hansson, 2007: 177). Unfortunately, as Schwandt (2015) acknowledges, “the processes by which synthesis judgments are made is not well understood, and there is little consensus in the field of evaluation about how to aggregate findings across multiple criteria or across stakeholders’ differing perspectives on important criteria” (p. 59). Scriven (1993) put it more bluntly; “Pulling it all together’ is where evaluators fall apart” (p. 72). A value-for-money



analysis often depends on synthesis. When value is represented in monetary units, it may appear that the problem has been solved because it has been reduced to counting money. However, monetary valuations may be incomplete because analysts tend to ignore aspects that are “just too hard to estimate” (Adler and Posner, 2006: 78). Moreover, monetary units may be a convenient way to measure value that is not economic in nature, in which case the units may have useful metric properties yet lack many of the properties associated with money and should not be aggregated with economic value without explicit justification (Gargani, 2017). Consequently, synthesis remains a problem even when value is monetized. Rubrics do not solve the synthesis problem entirely, but they make it possible to combine value of different types (Davidson, 2005).

## Evaluative rubrics with economic criteria

Evaluators may answer the question of value for money, like any evaluative question, using the *general logic of evaluation* (Scriven, 1991). The general logic is premised on a distinction between *description* (answering to the question “What’s so?”) and *valuing* (answering the question, “So what?”). Thus, evaluators describe what is being evaluated and learn from the people who are affected how much they value what has been described. With many descriptions and values in hand, evaluators *synthesize* them in a manner that warrants an *evaluative conclusion* of merit, worth, or significance.

There are many ways to put the general logic of evaluation into practice (Fournier, 1995; Schwandt, 2015), and rubrics are one (Davidson, 2005). They are relatively simple, apply evidence systematically, make the logic of evaluation transparent, are easily revised, and can be implemented in an inclusive, participatory manner. Rubrics are commonly organized as a matrix of criteria and standards. Criteria identify important characteristics of what is being evaluated, like the magnitude of a program’s impacts, its approach to inclusion, or its level of efficiency. Evaluators describe the characteristics identified by the criteria, and the standards establish how much value people place on what has been described. Typically, rubrics are constructed at the outset of an evaluation, for example, by working with stakeholders to understand how large an impact must be to matter to them or what types of inclusion they find desirable. Evaluators may express standards as levels of performance or success. Here, we describe them as levels of value. The levels may be binary (valuable or not), categorical (qualitatively different conceptions of value), ordinal (discrete levels of greater or less value), continuous (a continuum of better to worse), or a combination (such as a continuum with a binary cut off for minimum acceptable value). Once a rubric is constructed, evaluators gather evidence that allows them to compare a description of each criterion to its standard of value. Then evaluators synthesize the evidence across criteria to arrive at an overall evaluative judgment that answers the question, “To what level of value does the evidence point?”

A value-for-money rubric may include economic criteria, standards, and evidence. It should also include others that are noneconomic to judge resource use holistically in a given context (King, 2019; Schwandt, 2015). In practice, however, value-for-money studies frequently use narrow, decontextualized “defaults” provided by government agencies with a history of commissioning economic studies. A leading example is the Foreign, Commonwealth & Development Office framework, which has received a great deal of attention among development professionals. It sets out five generic criteria (economy, efficiency, effectiveness, cost-effectiveness, and equity, specifically defined) that the development agency finds

meaningful (Department for International Development (DFID), 2011). Stakeholders, however, may not find these concepts meaningful, which is why King et al. (2023) recommend defining contextual criteria which may include modifying, adding to, and removing from the generic criteria to construct rubrics that represent the interests of specific stakeholders in specific contexts.

### Example

To illustrate how context-specific criteria and standards may be applied to answer a value-for-money question, let us imagine that agricultural researchers in Nicaragua have developed a new variety of red bean they believe may have benefits related to:

- Sustainability because cultivation requires less water than current varieties;
- Nutrition because farmers will experience greater yields, reducing undernourishment and food insecurity;
- Equity because cultivation is well-suited to smallholder farms, many of which are owned by women who can sell beans to supplement their incomes;
- Culture because the beans may be used in traditional foods without altering their taste and appearance; and
- Economic efficiency because the cost of introducing the new variety is relatively low and the increase in yield is expected to provide benefits greater than the cost.

The researchers implemented a program to introduce the new variety to farmers. They secured funding and worked with local communities to develop a value-for-money rubric that included explicit criteria and standards that collectively define what “good resource use” means to stakeholders in this context (Table 1). For simplicity, our rubric only shows standards at one level, *good*. We would typically define standards for a range of levels that may include, for example, *excellent resource use*, *acceptable resource use*, and *poor resource use*. Also, for simplicity, we assume that the evaluator can measure and describe criteria to the level of precision necessary to distinguish *good* from *less than good*. In practice, we would ensure this is the case before proceeding and would provide justification for thresholds such as “less than 8% of community.” The rubrics below are for illustrative purposes only.

**Table 1.** A simple example of a rubric for answering the question “good resource use?”

Type of benefit	Criteria for good resource use	Standards for good resource use
Nutrition	1. Prevalence of undernourishment	Less than 8% of community
	2. Prevalence of food insecurity	Less than 12% of community
Culture	3. Beans are considered acceptable for traditional meals	No more than 5% find unacceptable
Equity	4. Income of women farmers	Average increase in income equivalent to US\$1500/year, bringing them close to parity with male farmers
Sustainability	5. Water stress	The ratio of local water demand to water supply is less than 70%
Economic Efficiency	6. Program cost	No more than US\$600 per farmer
	7. Social benefit-cost ratio	Greater than 1 (i.e. benefits exceed costs)



The criteria in Table 1 typically associated with value for money—total cost and benefit–cost ratio—are only two criteria out of seven. A narrow economic analysis might focus exclusively on these two to answer the question, “Is this good resource use?” But imagine the program were judged economically efficient in this narrow sense yet failed to achieve success on the other criteria. Would that constitute resources well used? Not in a substantive sense because investors and stakeholders are not getting what they paid for—a world that is better in specific ways.

Also note that equity, which is often peripheral to the economic analysis of benefits and costs, is included in the rubric. Here, it has an economic criterion and standard, yet equity is not solely an economic concept, and its inclusion may be motivated by a more comprehensive analysis of societal attitudes toward gender roles. This illustrates how a holistic value-for-money rubric may sometimes blur the line between what is economic and what is not. In the end, the distinction between economic and noneconomic is unimportant. What is important is crafting a rubric that adequately reflects what matters to people.

## Principles

Many of the funders, investors, policymakers, programs, and social enterprises with which we work start with a narrow interpretation of value for money. Sometimes it is sufficient to meet the needs of stakeholders, but often a more holistic approach is warranted. To help them expand their approach, we developed three principles to guide their evaluations.

### *Principle 1: Value depends on the credibility of evidence*

Imagine that two organizations implement the program described in the example. They work in separate communities that are identical in every way, and you want to invest in both. One organization conducted an evaluation you judge to be exemplary, and it provides evidence that the organization was successful on all criteria in Table 1. The second organization also conducted an evaluation, but you judge it to be poorly designed and executed. It provides evidence of identical results. Should you invest equally in the two programs?

The economically sound answer is no. You should provide more funding to the organization with better evidence, all else equal, because it imposes less *impact risk*. Impact risk is the risk that an organization will not create the impacts stakeholders desire and/or create impacts they do not desire (McLean and Gargani, 2019; see also Islam, 2023). When the evidence for impacts is more credible, we are more certain that impact estimates represent the true underlying impacts, thus impact risk is lower. The lower the impact risk, the more one should be willing to pay for the same impacts (all else equal). This is called *risk adjustment*.

Risk adjustment is a foundational concept in economics and finance for which there are many methods (Bernstein, 1996). A classic method is probability weighting (Bernoulli, 1954 (1738)). For example, if there is an 80 percent chance an investment will earn US\$100,000, the risk-adjusted earnings may be calculated as  $80\% \times \$100,000 = \$80,000$ . The need for risk adjustment extends beyond the domain of money to value in all its forms, and the general logic of evaluation helps us see why.

The general logic directs evaluators to describe the characteristics of what is being evaluated, such as a program’s impacts, then learn how much and what type of value stakeholders place on what was described. Importantly, stakeholders place value on

imperfect descriptions of reality (like impact estimates) not reality (the true underlying impact). So whatever uncertainty stakeholders may have about how much and what type of value they would place on something, it is compounded by their uncertainty about the true nature of what they are valuing. This is called propagation of error in statistics and the physical sciences (Young, 1962). From this, we can derive what might be considered a law of valuing—one cannot be more certain about value than the description of what is being valued (Gargani, 2016; Gargani and King, 2022).<sup>1</sup> An implication of the “law” is that *the value people place on an impact decreases as the credibility of its estimate decreases (all else equal)* (Gargani, 2015), which suggests that evaluators, like investors, should adjust value for risk.

How might this be done in a value-for-money evaluation? Let’s consider the equity criterion presented in Table 1. First, we can estimate the value stakeholders place on the program’s impact on equity without adjusting for risk. Let’s call this *unadjusted value*. One way to estimate it is

$$\text{Unadjusted Value} = \text{Impact Estimate} \times \text{Estimated Value per Unit of Impact} \quad (1)$$

For example, if an evaluator estimates that 10 women increased their yearly income because of the program (the impact estimate) and income increased US\$1500 on average (estimated value per impact), the unadjusted value would be  $10 \times \text{US\$1500} = \text{US\$15,000}$ . However, we are uncertain that precisely 10 women increased their incomes because of the program *and* that an impact is worth precisely US\$1500 on average to stakeholders. We can express our uncertainty about these estimates as certainty weights (which are different from importance weights). Let’s use the weight  $U$  to indicate how certain we are about the impact estimate and the weight  $W$  to indicate how certain we are about the value estimate. Let the weights range from 0 (the estimate is so uncertain we should ignore it) to 1 (the estimate is so certain we should treat it as the truth). From this, we can estimate the *adjusted value* as

$$\begin{aligned} \text{Adjusted Value} &= (U \times \text{Impact Estimate}) \times (W \times \text{Estimated Value per Unit of Impact}) \\ &= U \times W \times \text{Unadjusted Value} \end{aligned} \quad (2)$$

Equation (2) illustrates one approach among many to risk adjustment—reduce the value of the impact by the factor  $U \times W$ . Given that  $U$  and  $W$  range from 0 to 1 and can never be 1 (because no estimate is perfect), it follows that  $U \times W < U$ , which demonstrates the “law” and that the value people place on an impact decreases as the credibility of the impact estimate decreases, all else equal.

Estimating certainty weights, either subjectively or empirically, is a challenge and beyond the scope of this article. However, evaluators and stakeholders can, at the very least, construct subjective weights that reflect their judgments of certainty and credibility. The Intergovernmental Panel on Climate Change uses this approach to quantify the confidence experts place in evidence. Groups of experts express their confidence on a scale of *very high* to *very low*, taking into consideration the number, quality, and agreement of lines of evidence (Mastrandrea et al., 2011). The same could be done in value-for-money evaluations.

Table 2 provides an example. It revisits the scenario in which two organizations implement the Nicaraguan red bean program in separate but identical contexts. Program 1 was evaluated well and Program 2 poorly, and both evaluations produced identical findings as reflected in their unadjusted scores. Unadjusted sum scores were computed by assigning each criterion 1

**Table 2.** The value of identical programs (1 and 2) depends on certainty weights assigned to descriptive evidence (U) and valuations (W).

Criteria	Value	U	W	U × W × value
(a) The value of program 1 estimated with evidence from a well-conducted evaluation.				
1	1	0.9	0.9	0.8
2	1	0.9	0.9	0.8
3	1	0.9	0.9	0.8
4	1	0.9	0.9	0.8
5	1	0.9	0.9	0.8
6	1	0.9	0.9	0.8
7	1	0.9	0.9	0.8
Total value	Unadjusted = 7			Adjusted = 6
(b) The value of program 2 estimated with evidence from a poorly conducted evaluation. <sup>a</sup>				
1	1	0.5	0.9	0.4
2	1	0.5	0.9	0.4
3	1	0.5	0.9	0.4
4	1	0.5	0.9	0.4
5	1	0.5	0.9	0.4
6	1	0.5	0.9	0.4
7	1	0.5	0.9	0.4
Total value	Unadjusted = 7			Adjusted = 3

<sup>a</sup>The “round half to even” rule is applied.

if its supporting evidence indicated success and 0 otherwise, then summing the criterion scores. Both programs have unadjusted sum scores of 7, the maximum possible. However, we know that Program 2 is less valuable than Program 1 because it imposes greater impact risk. Subjective certainty weights allow us to estimate how much less by averaging the judgments of a group of experts and/or other stakeholders. Based on the analysis in Table 2, stakeholders should place about twice as much value on Program 1 than they do on Program 2 because Program 1 imposes less impact risk. Thus, Program 1 warrants greater investment (arguably about twice as much, assuming Program 2 is judged worthy of investment). The uncomfortable truth is that evaluators rarely make risk adjustments. Instead, we tend to treat evidence as if it is fully and equally informative, which suggests that we, as a field, may consistently overestimate value in general and value for money in particular. It is worth noting that the certainty of impact estimates may improve over time, for example, as more participants benefit or the benefits intensify, thereby making estimates less prone to error. This is one reason value-for-money assessments are repeated at regular intervals.

### *Principle 2: People place value on things*

The second principle is concerned with where value is located. It asserts that things do not have value, people place value on things (Gargani, 2018). Consequently, an evaluator cannot ascertain the value of an impact solely by studying it because value is not located within it. Value comes from people, who ascribe it to what they understand the impact to be.

**Table 3.** A value-for-money rubric that identifies whose value judgments are represented.

Type of benefit	Criteria for good resource use	Standards for good resource use	Whose value?
Nutrition	Prevalence of undernourishment	Less than 8% of community	Researchers
	Prevalence of food insecurity	Less than 12% of community	Researchers
Culture	Beans are considered acceptable for traditional meals	No more than 5% find unacceptable	Farmers
Equity	Income of women farmers	Average increase in income equivalent to US\$1500/year, bringing them close to parity with male farmers	Researchers
Sustainability	Water stress	The ratio of local water demand to water supply is less than 70%	Researchers
Economic Efficiency	Program cost	No more than US\$600 per farmer	Researchers
	Social benefit-cost ratio	Greater than 1 (i.e. benefits exceed costs)	Researchers

The implication is that evaluators must learn from people, especially those affected, how much and what type of value they place on impacts.

Some philosophers argue that certain entities, notably humans, other sentient creatures, and the natural world, have value inside of them. Value is “already there, discovered, not generated, by the valuer” (Rolston, 1987: 117). Others claim that this view of “intrinsic value as a mind-independent property is seriously flawed,” because, “human beings lack any evidence for this position and hence are unjustified in holding it” (Svoboda, 2011: 25). This remains a point of philosophical debate and will likely continue to be, but it has practical import for evaluators. Treating value as something that resides within people is consistent with evaluation’s guiding principles, standards, and competencies, which direct professional evaluators to understand and champion the diverse value perspectives of stakeholders (Yarbrough et al., 2010). Some people may value nature or something else because they believe it has intrinsic value, and evaluators should treat it as such when representing their perspective. However, this tells us what type of value people place on something, not where value is located. In the end, we conduct evaluations to be used *by people*. So regardless of where one believes value resides or should reside, it is the value people place on things that determines use.

We can put the second principle into practice by answering the question posed by Chambers (1995), “Whose reality counts?” The answer is found in Table 3 in the column labeled “Whose value?” For simplicity, we limit our consideration to two stakeholder groups, researchers and farmers. What we find is that six criteria and associated standards reflect the perspective of researchers; only one reflects the perspective of farmers. This type of information has the potential to change how people judge the legitimacy of the rubric, as it likely would in this case, and reminds evaluators they must justify their choices. In many real-world contexts, it would be difficult to defend this choice.

### *Principle 3: People value the same things differently*

Evaluators have long understood that people may not agree on which impacts matter, how much they matter, or why. In 1967, Suchman wrote, “From the point of view of evaluation,

conflicting values introduce serious problems for the determination of the criteria by which the success of a public service program is to be judged” (p. 37). He offered examples from the 1950s in which various groups used either economic costs or social consequences as evaluative criteria. The choice of criteria mattered because, he observed, “a public service program may be judged desirable or successful by one scheme of values and undesirable or unsuccessful according to another.” Is this really a problem?

We and other contemporary evaluators do not believe it is. Instead, we believe that variation in value perspectives is information; evaluators have a duty to ascertain, understand, and report this information; and when value perspectives conflict, evaluators should help reconcile them, at least within the context of conducting an evaluation (Gargani, 2018; Gargani and King, 2022). This is consistent with *The Program Evaluation Standards* (Yarbrough et al., 2010), which direct evaluators to:

- “Learn what stakeholders value about the program, how strong these values are held, and the degree to which these values converge or conflict” (Standard U4, p. 39);
- Refrain from “assuming the values of decision makers reflect those of the program personnel and program users” (Standard U2, p. 25); and
- Justify evaluation conclusions and decisions in a way that takes diverse cultures, contexts, and values into account (Standard A1, pp. 165–167).

Similar guidance is given to evaluators around the world in documents such as *The 2018 AEA Evaluator Competencies* (King and Stevahn, 2020, Competencies 3.2, 3.7, and 5.2), OECD’s (2021: 41, 65) *Applying Evaluation Criteria Thoughtfully*, the *Evaluation Standards for Aotearoa New Zealand* (Aotearoa New Zealand Evaluation Association and Social Policy Evaluation and Research Unit, 2015), and the Japan Evaluation Society’s (2012) *Guidelines for the Ethical Conduct of Evaluation*. Despite this consensus, evaluators are often expected to reach a single, overarching evaluative conclusion—even when stakeholders disagree about what matters. When this is the case, a single evaluative conclusion may be inadequate, and evaluators should reach multiple, possibly conflicting conclusions about value and value for money that reflect all perspectives (Gargani, 2017).

Rubrics can help. Table 4 adds two new columns to the example rubric, one for researchers and the other for farmers. In the context of cost–benefit analysis, separating stakeholders (sites, funders, or other meaningful groups) like this is referred to as keeping *multiple accounts* of benefits and/or costs (Shaffer, 2010). We have adapted multiple accounts for rubrics to express different value perspectives. In general, we can do this in four ways.

**Stakeholder groups apply different standards.** In Table 4, researchers require less than 8 percent of the community to be undernourished to judge the result valuable, while farmers would judge rates as high as 20 percent to be valuable. So, it is possible that farmers will value the nutritional impact achieved by the program and researchers will not.

**Stakeholders seek different evidence.** This is the case for sustainability in Table 4. Researchers favor a standard based on the ratio of local water demand to supply, and farmers favor a standard based on farmers having as much water as they want for cultivating red beans. Again, one group may place value on the impact that the program achieves and the other may not.

**Table 4.** A multiple account rubric for answering the question “good resource use?”

Type of benefit	Criteria	Standards	
		Researchers	Farmers
Nutrition	Prevalence of undernourishment	Less than 8% of community	Less than 20% of community
Culture	Prevalence of food insecurity	Less than 12% of community	Less than 10% of community
	Beans are considered acceptable for traditional meals	No more than 5% find unacceptable	0% find unacceptable
Equity	Income of women farmers	Average increase in income equivalent to US\$1500/year, bringing them close to parity with male farmers	Average increase in income equivalent to US\$2000/year, bringing them close to parity with male farmers
Sustainability	Water stress	The ratio of local water demand to water supply is less than 70%	90% of farmers have as much water as they want to cultivate beans
Economic Efficiency	Program cost	No more than US\$600 per farmer	Not important
	Social benefit-cost ratio	Greater than 1	Greater than 2

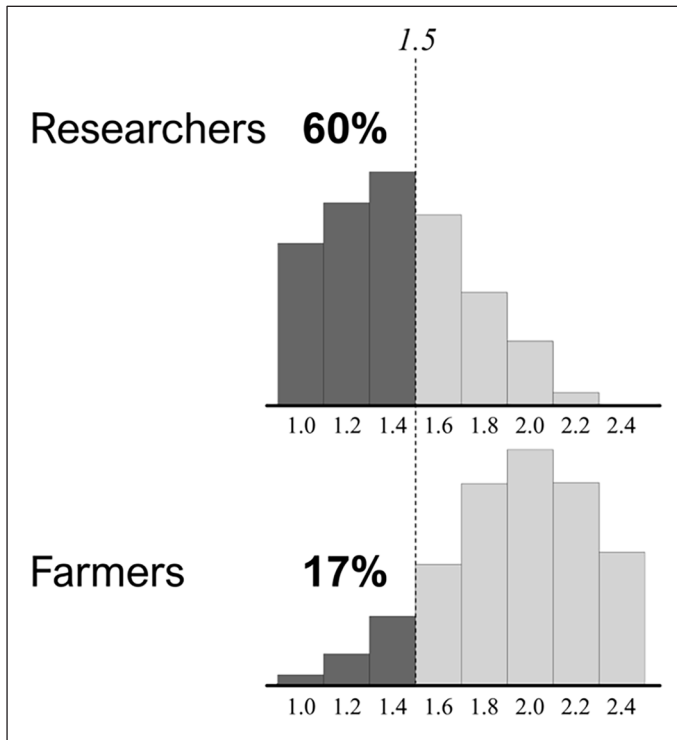
**Stakeholders use different criteria.** In Table 4, farmers do not find the criterion of total cost important, and researchers do. This is the equivalent of farmers assigning the criterion an importance weight of 0. This may lead the groups to different evaluative conclusions.

**Stakeholder groups apply different importance weights.** For example, instead of farmers assigning total cost an importance weight of 0, they could assign a weight of 0.5. This would indicate they value it half as much as researchers do.

Evaluators can go further to understand and express variation in value perspectives. Within accounts, it is unlikely that agreed-upon criteria, standards, and measures precisely match those preferred by individuals. Rubrics are often developed through a process of participation, facilitation, and negotiation, and their final structure reflects a compromise that is acceptable to those involved. The structures preferred by individuals, however, may still vary substantially, so evaluators may wish to track individual variation and estimate the proportion of stakeholders within each group who judge a result valuable using each individual’s combination of weights, criteria, standards, and evidence.

Figure 3 illustrates how this may be done using one criterion from the example—the social benefit–cost ratio. Researchers agreed that a ratio greater than 1.0 would be judged valuable, and farmers agreed that a ratio of 2.0 would be judged valuable. If the estimated ratio were 1.5, an evaluator would conclude that researchers were satisfied, and farmers were not. Within each group, however, individuals preferred different cut offs, some higher and some lower, making it unlikely all group members would be satisfied (or not) with the result. Figure 3 shows that the consensus among researchers was low. Even though the estimated social benefit–cost ratio is 1.5, higher than their agreed-upon standard of 1.0, 40 percent of researchers are not satisfied with the result. Farmers were more tightly clustered around their negotiated standard of 2.0, thus only 17 percent are satisfied with the result of 1.5. Similar differences may be found on other criteria, bringing the validity of a single evaluative conclusion into question.





**Figure 3.** Histograms for researchers and farmers display (1) variation in what individuals judge to be a valuable social benefit-cost ratio and (2) group differences in the percentage of stakeholders satisfied with the estimated ratio of 1.5.

## Conclusion

Value for money poses an evaluative question about resource use at the intersection of evaluation and economics. To answer the question well, evaluators should draw on the theory and practice of both disciplines, particularly the general logic of evaluation and the use of explicit criteria to describe what people genuinely value. Value-for-money criteria may include some that are economic (like financial costs, monetary valuations of benefits, and benefit–cost ratios) alongside others that economic analysis struggles to address (like equity, social justice, and inclusion). Using familiar evaluation tools like rubrics, evaluators can conduct holistic value-for-money assessments as they would other evaluations.

That is not to say that the question of value for money is easy to answer. While it is rooted in simple notions of fairness and getting what one pays for, the principles we introduced highlight some challenges. People should place less value on an impact as the credibility of its estimate decreases, posing the challenge of adjusting value for the associated risk. Understanding what people value, how much they value it, and the kind of value it has for them poses additional challenges. Finally, recognizing that people value things differently, often with substantial variation, presents the challenge of reconciling and reporting diverse evaluative conclusions.

By addressing these challenges, we not only enhance our understanding of value for money but may also ease, if only a little, the long-standing synthesis problem. The more heterogeneous an evaluation's criteria, standards, measures, and stakeholders, the more difficult the synthesis. By constructing rubrics with multiple accounts, as we did for different stakeholder groups, we segment variation. The heterogeneity of value perspectives within accounts may be less than the heterogeneity across all stakeholders when the value perspectives of stakeholder groups differ. Thus, syntheses within accounts may be more tractable than it would be across all stakeholders when evaluators may draw separate value-for-money conclusions for groups and individuals.

Multiple, sometimes conflicting conclusions provide valuable insights about resource allocation. Although it may not always be feasible, why is it not the norm? Multiple conclusions, especially those that contradict or conflict, are what is demanded by evaluation standards. Moreover, it is precisely the information funders, investors, entrepreneurs, and managers need to allocate resources well.

Rubrics offer a practical tool for managing the complexities of synthesizing multiple criteria, standards, lines of evidence, and their associated variation and uncertainty. A value-for-money rubric can incorporate both economic and non-economic criteria, standards, and evidence, and can blend both technocratic and democratic approaches to decision-making. From a technocratic perspective, a rubric is a matrix of criteria and standards and a tool for making a set of values explicit. From a democratic perspective, a rubric supports power sharing to co-create context-specific definitions of value for money with stakeholders and to make evaluative judgments systematically and transparently. Used in this way, rubrics support holistic assessments tailored to specific contexts and stakeholder interests.

The use of rubrics in value-for-money assessment is in its nascency, with significant room for further development in its application. We have demonstrated that value-for-money rubrics are able to incorporate risk adjustment, make transparent whose values are represented, and keep the values of different groups separate and distinct. The fact that this is not routinely done highlights an opportunity to improve evaluations of resource use.

## Note

1. There is a "law of propagation of uncertainty" in the physical sciences that is not a law, per se, but a mathematical technique for estimating the effect of combined errors.

## References

- Adler MD and Posner EA (2006) *New Foundations of Cost-Benefit Analysis*. Cambridge, MA: Harvard University Press.
- Aotearoa New Zealand Evaluation Association and Social Policy Evaluation and Research Unit (2015) Evaluation standards for Aotearoa New Zealand. Available at: <https://thehub.swa.govt.nz/assets/Uploads/Superu-Evaluation-standards.pdf>
- Belcher B and Palenberg M (2018) Outcomes and impacts of development interventions: Toward conceptual clarity. *American Journal of Evaluation* 39(4): 478–95.
- Bernoulli D (1954 [1738]) Exposition of a new theory on the measurement of risk (trans. L Sommer). *Econometrica* 22(1): 23–36.
- Bernstein PL (1996) *Against the Gods: The Remarkable Story of Risk*. New York: Wiley.
- Brookings (2023) Global impact bond database snaps October 1, 2023 for the ethical conduct. Available at: <https://www.brookings.edu/wp-content/uploads/2023/08/Impact-Bonds-Snapshot-October-2023.pdf>

- Chambers R (1995) Poverty and livelihoods: Whose reality counts? *Environment and Urbanization* 7(1): 173–204.
- Davidson EJ (2005) *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: Sage.
- Department for International Development (2011) *DFID's Approach to Value for Money (VfM)*. London: DFID, UK Government.
- Destremau K and Wilson P (2017) Defining social investment, Kiwi-style. In: Boston J and Gill D (eds) *Social Investment: A New Zealand Policy Experiment*. Wellington, New Zealand: Bridget Williams Books, 32–79.
- Dodgson JS, Spackman M, Pearman A, et al. (2009) *Multi-Criteria Decision Analysis: A Manual*. London: Department for Communities and Local Government.
- European Commission (1999) *MEANS Collection: Evaluating Socio-Economic Programmes*. Vol. 4. Technical solutions for evaluation within a partnership framework. Luxembourg: European Commission.
- Flyvbjerg B and Bester D (2021) The cost-benefit fallacy: Why cost-benefit analysis is broken and how to fix it. *Journal of Benefit-Cost Analysis* 12(3): 345–419.
- Fournier DM (1995) Establishing evaluative conclusions: A distinction between general and working logic. In: Fournier DM (ed.) *Reasoning in Evaluation: Inferential Links and Leaps. New Directions for Evaluation*. Vol. 58. Hoboken, NJ: Jossey Bass, 15–32.
- Fraser A, Tan S, Lagarde M, et al. (2018) Narratives of promise, narratives of caution: A review of the literature on social impact bonds. *Social Policy & Administration* 52(1): 4–28.
- Gargani J (2015) Valuation: Methods for representing the value of program impacts in monetary units. In: *The annual conference of the Canadian Evaluation Society*, Montreal, Canada, 24–27 May.
- Gargani J (2016) An introduction to social return on investment (SROI) [Workshop]. In: *The annual conference of the Canadian Evaluation Society*, St. John's, Canada, 5–8 June.
- Gargani J (2017) The leap from ROI to SROI: Farther than expected? *Evaluation and Program Planning* 64: 116–26.
- Gargani J (2018) Reflections on social impact measurement and management [Keynote address]. In: *The Social Impact Management Initiative (SIMI) Social Impact Day*, Tokyo, Japan, June.
- Gargani J (2019) *Insights from John Gargani. Impact Responsibility*. San Francisco, CA: Salesforce, 72–74.
- Gargani J and King J (2022) Disrupting value for money: New principles and methods (we hope you will find challenging). In: *The biennial conference of the European Evaluation Society*, Copenhagen, Denmark, 6–10 June.
- Hansson SO (2007) Philosophical problems in cost–benefit analysis. *Economics & Philosophy* 23(2): 163–183.
- Holland PW (1986) Statistical and causal inference. *Journal of the American Statistical Association* 81(396): 945–60.
- Islam SM (2023) Impact risk management in impact investing: How impact investing organizations adopt control mechanisms to manage their impact risk. *Journal of Management Accounting Research* 35(2): 115–139.
- Japan Evaluation Society (2012) *Guidelines for the Ethical Conduct of Evaluation (Unofficial English Translation)*. [http://evaluationjp.org/files/Guidelines\\_for\\_Ethics\\_20131106.pdf](http://evaluationjp.org/files/Guidelines_for_Ethics_20131106.pdf)
- Julnes G (2012) Managing valuation. In: Julnes G (ed.) *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation: New Directions for Evaluation*. Vol. 133. Hoboken, NJ: John Wiley & Sons, 3–15.
- Kahneman D (2012) *Thinking Fast and Slow*. London: Penguin Press.
- Keeney RL and Raiffa H (1976) *Decisions with Multiple Objectives: Performances and Value Trade-Offs*. New York: Wiley.
- King J (2017) Using economic methods evaluatively. *American Journal of Evaluation* 38(1): 101–13.
- King J (2019) *Evaluation and value for money: Development of an approach using explicit evaluative reasoning*. Doctoral Dissertation, University of Melbourne, Melbourne, VIC, Australia.

- King J (2023) How should program evaluation standards inform the use of cost-benefit analysis in evaluation? *Journal of Multidisciplinary Evaluation* 19(43): 34–48. Available at: [https://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/829](https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/829)
- King JA and Stevahn L (2020) Presenting the 2018 AEA evaluator competencies. *New Directions for Evaluation* 2020(168): 49–61.
- King J, Wate D, Namukasa E, et al. (2023) *Assessing Value for Money: The Oxford Policy Management Approach* (2nd ed.). Oxford: Oxford Policy Management Ltd. Retrieved from: <https://www.opml.co.uk/files/Publications/opm-vfm-approach-2.pdf?noredirect=1>
- Klonschinski A (2014) “Economic imperialism” in health care resource allocation—how can equity considerations be incorporated into economic evaluation? *Journal of Economic Methodology* 21(2): 158–74.
- Laplume AO, Sonpar K and Litz RA (2008) Stakeholder theory: Reviewing a theory that moves us. *Journal of Management* 34(6): 1152–89.
- McKevitt D (2015) Debate: Value for money—in search of a definition. *Public Money & Management* 35(2): 99–100.
- McLean R and Gargani J (2019) *Scaling Impact: Innovation for the Public Good*. London: Routledge.
- Martens KSR (2018) Rubrics in program evaluation. *Evaluation Journal of Australasia* 18(1): 21–44.
- Mastrandrea MD, Field CB, Stocker TF, et al. (2011) The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change* 108: 675–91.
- Organisation for Economic Co-operation and Development (2021) *Applying Evaluation Criteria Thoughtfully* (OECD). Paris: Organization for Economic Co-operation and Development.
- Rolston H (1987) *Environmental Ethics: Duties to and Values in the Natural World*. Philadelphia, PA: Temple University Press.
- Saul J, Rabinowitz D, Hussain I, et al. (2023) Impact markets: The next frontier. *Stanford Social Innovation Review*, 7 September. Available at: <https://doi.org/10.48558/JZG7-0930>
- Schwandt T (2015) *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*. Redwood City, CA: Stanford University Press.
- Scriven M (1991) *Evaluation Thesaurus*. Thousand Oaks, CA: Sage.
- Scriven M (1993) Intermediate evaluation design issues. *New Directions for Program Evaluation* 1993(58): 67–74.
- Scriven M (1994) The final synthesis. *Evaluation Practice* 15(3): 367–82.
- Scriven M and Davidson EJ (2000) The synthesis problem: Issues and methods in the combination of evaluation results into overall evaluative conclusions. In: *Skill-building demonstration presented at the meeting of the American Evaluation Association*, Honolulu, HI, November. Available at: <https://realevaluation.com/pres/SynthHNL.pdf>
- Shaffer M (2010) *Multiple Account Benefit-Cost Analysis: A Practical Guide for the Systematic Evaluation of Project and Policy Alternatives*. Toronto, ON, Canada: University of Toronto Press.
- Stern E, Stame N, Mayne J, et al. (2012) Broadening the range of designs and methods for impact evaluations. Available at: <https://www.oecd.org/derec/50399683.pdf>
- Svoboda T (2011) Why there is no evidence for the intrinsic value of non-humans. *Ethics and the Environment* 16(2): 25.
- UK Government (2020) *Guide to Using the Social Value Model*. London: Government Commercial Function. Available at: <https://assets.publishing.service.gov.uk/media/5fc8b804d3bf7f7f53e5a503/Guide-to-using-the-Social-Value-Model-Edn-1.1-3-Dec-20.pdf>
- White H (2010) A contribution to current debates in impact evaluation. *Evaluation* 16(2): 153–64.
- Yarbrough DB, Shulha LM, Hopson RK, et al. (2010) *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, 3rd edn. Thousand Oaks, CA: Sage.
- Young HD (1962) *Statistical Treatment of Experimental Data*. New York: McGraw-Hill.

---

John Gargani conducts research on evaluation methods related to impact, value, scaling, value for money, contingent payment for impacts, and risk-adjusted pricing of impacts.

Julian King's research interests center on evaluative reasoning, economic methods of evaluation, and interdisciplinary approaches to evaluation and value for money.