

The Barriers to Aggregating Heterogeneous Impact Indicators

The Case of Social Impact

White Paper

Author: Axelle Guers

Abstract

Impact measurement in impact investing remains a methodological field still under construction. On certain environmental topics, such as climate, established and shared metrics already exist — tonnes of CO₂ induced, avoided, or sequestered, for instance, now benefit from accounting standards and norms that make comparison and consolidation easier. Social impact, by contrast, offers a striking illustration of a more general difficulty: that of comparing and consolidating heterogeneous indicators. Its indicators are often granular, specific to each company's activities, and difficult to compare with one another. In practice, investors manage to consolidate volume indicators at fund level — number of beneficiaries reached, jobs supported — which are relatively homogeneous; but the indicators that capture the changes generated for stakeholders remain specific to each company and difficult to aggregate. It is precisely on these latter indicators that comparison is most difficult, even though they constitute the richest information on what a company's intervention actually produces.

This paper explores why comparing and consolidating heterogeneous indicators constitutes a structural challenge for impact investors and identifies the underlying mechanisms behind it. To examine this challenge, we focus on the case of social impact indicators, which are particularly illustrative on this difficulty.

We identify five main barriers. First, the granular and company-specific nature of social indicators gives rise to (1) an absence of impact objectives set according to comparable principles — comparing the achievement rate of an unambitious target to that of a more demanding target set on different principles is difficult — and to (2) a difficulty in ranking heterogeneous impacts against one another — does an improvement in health outweigh an improvement in foreign-language proficiency? In addition, (3) the heterogeneity of data — variable methodologies and degrees of data reliability — raises a question of comparability of measurement results: two seemingly similar values can rest on very unequal foundations. Next comes the question of (4) the fungibility of social impacts: aggregating impacts and establishing their substitutability requires normative trade-offs that cannot be resolved on technical grounds alone. Finally, when comparing two companies, (5) the heterogeneity of measurement scopes means that observed gaps may reflect framing choices as much as genuine differences in impact. These last three barriers also apply to indicators sharing the same unit, but it is their accumulation across heterogeneous indicators that creates the central difficulty.

In conclusion, we outline several avenues for addressing these barriers: ranking impacts and beneficiary vulnerability, defining objectives according to shared principles, harmonizing data-collection methodologies, and making measurement scopes explicit. All of these avenues call for greater transparency regarding the assumptions underlying social impact measurement.

Introduction

Impact investing refers to an investment approach that deliberately seeks to generate both a financial return and a positive, measurable social or environmental impact (World Economic Forum, 2013). It differs from socially responsible investing, which primarily aims to reduce extra-financial risks (Höchstädter & Scheck, 2015), through three specific pillars: the investor's ex ante intentionality, the additionality of its action, and the measurement of the social and environmental externalities generated (France Invest & FIR, 2021).

The targeted impacts are diverse, all the more so in the social domain that is the focus of this white paper: job creation in underserved areas, improved access to education, healthcare, or energy for underserved populations (France Invest & FIR, 2021). This diversity is inherent to the field and is visible among the companies financed: the same fund may simultaneously finance and support a company offering multidisciplinary healthcare for women in areas with insufficient specialized medical provision, a digital platform for career guidance and skills development accessible to beneficiaries around the world, including in countries of the Global South, and a company running prevention and social activities to combat isolation among older people. Each measures its impact using indicators specific to its own activity — reduced time to access care and patient satisfaction for the first, growing self-confidence in career progression for the second, and the proportion of beneficiaries reporting feeling less isolated for the third — and these indicators are fundamentally different from one another.

This white paper focuses precisely on this scenario: the comparison and consolidation of heterogeneous indicators, built for fundamentally different activities and missions. The difficulty raised by this heterogeneity is not specific to the social field — it confronts any investor seeking to report on its impact in aggregate form, particularly when managing a multi-sector portfolio. Social impact, however, is a particularly well-suited case study, given how diverse the effects generated for stakeholders can be.

Definition

We should specify here that, for the purposes of this white paper, we define social impact as the effects generated by an intervention on its stakeholders.

The literature distinguishes three levels of indicators:

- Output indicators: the products or services directly delivered by the company's activity — goods distributed, people trained, beneficiaries reached, etc.
- Outcome indicators: the medium- and long-term changes experienced by beneficiaries following the intervention.
- Impact indicators: significant and lasting changes to root causes, or at the level of communities or society (Ebrahim & Rangan, 2014).

In this paper, we focus mainly on outcome and impact indicators, which we refer to collectively using the term “social impact indicator.” This term captures a dimension of the quality and nature of the effect produced, not merely its volume, unlike output measurement.

As Kaufmann et al. (2025) point out, impact measurement has progressively become an essential characteristic of actors in the field, to the point of being among the practices now expected in order to operate and be recognized within the impact investing ecosystem — and this, even before coherent methodologies had been established in the social domain. The literature on impact evaluation, although long-standing — its origins trace back to the aftermath of the Second World War, in the context of the rise of social public policy and later the structuring of development aid (Sibieude et al., 2021) — remains fragmented and made up of heterogeneous methodologies that were not designed for the specific challenges of impact investing (Corvo et al., 2021).

This situation creates a particular tension for investors, who face two simultaneous requirements: managing the impact of portfolio companies as closely as possible to their operational mission, and reporting on that impact in a synthesized form to their stakeholders (Arjaliès et al., 2018).

In the absence of common standards, the objectives of transparency, comparability, and improved social performance therefore remain difficult to achieve. This observation calls for an examination of the underlying causes of this disconnect. This white paper explores the structural barriers to comparing and consolidating social impacts.

Methodology

This white paper draws on the academic literature on social impact measurement and impact investing, as well as on field observations conducted within Mirova, over a period of sixteen months, complemented by nineteen semi-structured interviews with impact investing professionals across various types of organizations — asset managers, institutional investors, companies, and consulting firms. The verbatim quotes and examples cited in this document are drawn from these interviews and field observations and are intended to illustrate and complement certain elements found in the literature.

1. In impact investing: a need to compare and consolidate impacts

Impact-focused asset management companies face three distinct objectives (Arjaliès et al., 2018).

The first is **impact management** for the companies supported. The investor tracks the progress of each company against its impact objectives, which requires having indicators that are clear and suited to its activities.

The second is **reporting** to investors. Asset management companies act as financial intermediaries and must synthesize impact-related information in order to report to their stakeholders. This requirement calls for indicators that are not only consolidated and communicable at portfolio level, but also sufficiently clear and comparable from one asset management company to another (Corvo et al., 2021).

*“With a social portfolio, we’ll have individual indicators, but what story do we tell at portfolio level? The idea is to have a few headline figures.” — Asset manager,
Impact expert*

”

The third is **allocation trade-offs**: investors also need to compare companies with one another in order to make trade-offs at fund level while taking impact into account.

*“You have €1 to invest — where do you invest it? The question we have to ask ourselves is how to maximize our impact as a funder. How do you rank these companies based on what they create, per euro invested?” — Asset manager,
Impact expert*

”

In practice, the consolidation of social indicators at fund level today relies mainly on volume indicators — number of beneficiaries reached, jobs supported — because these apply across a wide range of activities. But volume alone does not capture the quality of the impact or the full social value created: access to safe drinking water for rural households lacking basic infrastructure is not directly comparable to vegetarian cooking classes for urban dwellers concerned about food issues. Similarly, a literacy program for adults facing illiteracy cannot be directly compared to a foreign-language learning app for high-school students. Volume alone cannot capture these differences.

It is precisely this encounter between the necessary specificity of indicators at company level and the need for synthesis at fund level that creates the central tension explored in this white paper.

2. Comparing two impacts: the barriers at work

2.1. The absence of comparable objectives and the difficulty of ranking

This section highlights two barriers, both of which stem from the granular and company-specific nature of social indicators: **the absence of impact objectives set in a comparable way**, and **the difficulty of ranking heterogeneous impacts against one another**. To understand the origin of these two barriers, we must first return to the conditions that would, in theory, allow the social value created by different companies to be compared.

Ideally, to compare the social value created by different companies, one would need three elements (Kroeger & Weber, 2014):

Impact dimensions comparable in importance. For the comparison between two interventions to be valid, they must be assessed on dimensions that carry equivalent weight for beneficiaries' well-being. This is the principle put forward by Kroeger & Weber (2014), who propose measuring beneficiary satisfaction across different areas of life — health, education, social connection — and ensuring that the areas selected are sufficiently close in importance.

Objectives defined in a comparable way. For the comparison between two interventions to be meaningful, it must be possible to position their results relative to a shared reference point — a threshold defining what constitutes a situation of need, against which one can measure whether, and by how much, the gap has been reduced. Without this shared threshold, one can describe what an intervention produces in absolute terms but cannot assess its effectiveness or compare it to that of another intervention operating in a different context.

To this end, Kroeger & Weber (2014) propose using the national average satisfaction level in a given area of life — health, education, social connection — as a reference: individuals below this average are considered to be in a situation of need, and the degree of social value created is measured by the reduction of this gap. This approach makes it possible, in particular, to compare the effectiveness of two interventions by relating their results to objectives set consistently with their context: if a population is harder to reach, the reference threshold will naturally reflect this, making the comparison more meaningful. Kroeger & Weber propose this logic at the national level, but it could in principle be applied at a finer level of granularity — a region, a specific population.

This logic is already at work in certain impact investing practices. For indicators such as a living income or a living wage, investors commonly use geographic benchmarks, such as those produced by the Anker Research Institute. The target level varies by geography, but this is precisely what makes the comparison meaningful: what is being compared is not absolute levels, but gaps relative to a threshold grounded in local realities.

More broadly, the challenge is to have objectives defined according to a shared logic, making it possible to compare levels of achievement across interventions operating in different environments.

A counterfactual that isolates the intervention's own effect from what would have happened without it — without one, it is difficult to distinguish what is due to the company's action from what results from external factors such as economic conditions, ongoing public policies, or dynamics specific to the targeted population.

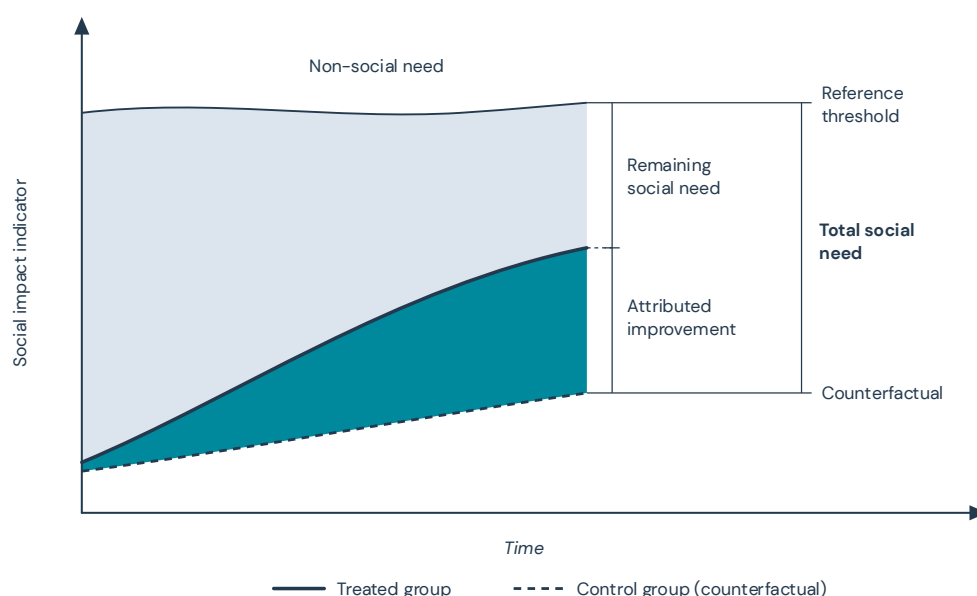


Figure 1. Framework for comparing social value created (adapted from Kroeger & Weber, 2014).

In this framework, the degree of value created corresponds to the attributed improvement divided by total social need.

These three conditions are rarely met in the context of impact investing.

2.1.1 Multiple causal chains and the difficulty of the counterfactual

Before formalizing the two resulting barriers (section 2.1.3), it is worth returning to their origin: the difficulty of the counterfactual and the use of indicators specific to each company's activities.

In the social domain, observed results rarely stem from the action of a single actor: several organizations, public policies, and contextual dynamics contribute simultaneously to the same outcome, making it very difficult to isolate the specific contribution of a given company. This is precisely the role of the counterfactual: by comparing the treated group to a control group with similar characteristics, it isolates the intervention's own effect from external factors.

Field observations

The ideal counterfactual is structurally difficult to construct in the context of impact investing. Collecting data from a control group requires significant resources, on top of the data collection already needed for direct beneficiaries and across an entire portfolio of companies. Annual reporting cycles leave little room for heavy,

frequent data collection. In practice, investors rely on approximations such as the self-reported counterfactual, which captures what the beneficiary attributes to the intervention.

“People who aren’t familiar with it often say, ‘Why don’t you just set up a comparison group?’ We’re working in a social, multifactorial setting — building a counterfactual is a heavy, costly method that isn’t well suited to this environment. A somewhat less robust method is the ‘before/after’ approach, but that too isn’t always suited given the resources available and the lack of forward planning. We favor pragmatic approaches that help the organization answer the evaluative questions it has and understand and measure its impact while trying to bring rigor throughout the process to avoid bias. So we make heavy use of mixed qualitative and quantitative methods, often ex post. Of course, we distinguish between how someone feels right after the fact, which carries biases, and how they feel after 3 or 6 months. We then ask questions designed to attribute the changes: ‘Do you feel you’ve developed these skills? Is that thanks to the program?’” — Consulting firm, Impact expert

This approach, however, has an important limitation: the relevance of the answer depends closely on the precision of the question asked. If a participant in a one-week professional training course is asked how much their general level of education has improved, the question is too broad to be meaningful — the person may struggle to isolate the effect of that training from their overall path. By contrast, if the question focuses precisely on the skills targeted by the training — for example, their ability to write a CV or conduct an interview — the answer becomes more reliable and more clearly attributable to the intervention. The quality of the self-reported counterfactual therefore depends on the granularity and precision of the questions asked.

Ebrahim & Rangan (2014) formalize this difficulty and show, based on two cases they studied, that causal attribution is only credible under two conditions. Either the link between the organization’s action and the outcome is very well documented scientifically — as with the Aravind Eye Hospital in India, where cataract surgery directly and measurably restores patients’ vision. Or the organization controls the entire intervention pathway needed to produce the outcome — as with the Harlem Children’s Zone in New York, which supports children from a disadvantaged neighborhood from birth through to entry into employment, integrating education, family support, health, and job placement within a defined geographic area.

These two conditions are exceptional. For the vast majority of impact companies, final outcomes result from multiple contributions that no single actor controls.

Field observations – The impact investor: an aggregator that does not control the chain

Ebrahim & Rangan suggest that it is funders — foundations, governments, development agencies — who are best positioned to measure systemic impacts, since they have an overview of multiple organizations contributing to the same objective.

This logic does not translate directly to impact investors. Some funds are structured around a shared theme — health, employment integration, access to energy — which creates overall coherence. However, the companies involved often operate across varied geographies, serving distinct populations, and are not necessarily part of a coordinated intervention pathway within a defined territory. Some funds also adopt a multi-sector approach, which moves them even further away from the integrated-ecosystem logic described by Ebrahim & Rangan.

Impact investors thus occupy an intermediate position: too close to individual companies to have a systemic view of causal chains, yet required to compare them with one another to inform their investment decisions.

2.1.2 Toward indicators specific to companies' activities

To address the problem of comparability of social impact indicators, several authors have proposed converging toward a common core of standardized indicators (Ferreiro & Kölbel, 2025; Crucke & Decramer, 2016). As early as 2014, the Comité Français sur l'investissement à impact social called for a shared measurement framework. Standardization would offer a further advantage: limiting the selection bias that leads to retaining only the most favorable impacts (Arjaliès et al., 2018; Kaufmann et al., 2025).

Ebrahim & Rangan's (2014) analysis argues, however, that companies should focus on measuring what they do directly, in connection with their operational mission, rather than seeking to claim credit for outcomes arising from broader causal chains.

Field observations

While the academic literature framed this logic mainly at the company level, our observations show that it also applies at the investor level: investors often select, and frequently co-construct with companies, indicators anchored in their direct activities rather than in final outcomes that partly lie beyond their control.

Take the example of an online career-guidance platform that helps learners clarify their goals and identify possible paths. Returning to employment may then depend on additional training, support from other organizations, or local labor-market conditions — all factors beyond the platform's control. Its intervention constitutes an intermediate step toward employment but cannot alone guarantee that outcome. A standard indicator such as return to employment therefore appears poorly suited to assessing this company's

impact: on this point, one impact expert highlighted “the difficulty of demonstrating impact and causality, given that the intervention is only one step among others on the path to employment.” Advisory experts therefore recommended measuring instead the improvement in learners’ self-confidence regarding their career path — a level at which the company can claim a contribution that is easier to isolate.

This logic is reinforced by the literature on social impact measurement: given the diversity of impact companies — heterogeneous sectors, different geographies, varied intentionalities (Höchstädter & Scheck, 2015) — measurement must be tailored to the size, sector, and degree of social innovation of each organization (Mulgan, 2010; Nicholls, 2009), and must enable effective management closely tied to companies’ concrete activities (Leclerc & Sibieude, 2025).

This logic, however, comes at a direct cost to the comparability of measurements.

2.1.3 Two cascading barriers: the absence of comparable objectives and the difficulty of ranking

The reliance on company-specific indicators produces two barriers to the comparability and consolidation of social impacts. First, the absence of a standardized indicator makes it difficult to define a reference level established on a shared principle. Kroeger & Weber (2014), for instance, propose using as a reference threshold the average satisfaction already achieved by other individuals in a given region for a given area of life. The threshold thus remains local, but it rests on a comparative principle: what others have achieved. This approach has the advantage of indirectly capturing the difficulty of the context — in a region where satisfaction levels are structurally low, a modest improvement represents relatively significant progress. Living-wage or living-income methodologies rest on a different principle: the threshold is likewise defined locally, but normatively, based on the resources deemed necessary for a decent standard of living, independent of the level achieved by others. This solution breaks down once indicators become company-specific: there is then neither an established normative benchmark nor an available comparative level to define, for example, what an ideal level of self-confidence regarding one’s career path would look like.

Field observations

In practice, investors sometimes define impact targets at the level of each company, which makes it possible to measure an aggregate target-achievement rate. But these targets are often set case by case, notably based on historical results and assumptions used to project them into the future. These assumptions are themselves more or less conservative or ambitious, with no shared external principle to frame them — resulting in requirements that vary from one company to another.

Comparing an 80% achievement rate based on an easy target to a 60% rate based on an ambitious one tells us nothing about the relative social value produced.

Second, even if two companies achieve good levels on their respective indicators, it is difficult to rank their impacts without a shared value framework. Where Kroeger & Weber worked with generic satisfaction domains whose relative importance is assumed to be comparable, here we are instead dealing with highly heterogeneous indicators whose relative importance has not been established. Does a satisfactory improvement in health outweigh a satisfactory improvement in foreign-language proficiency? The question has no technical answer — it is a normative one.

This difficulty has another dimension: the comparability of indicators and the comparability of beneficiaries are linked. If two needs are comparable in nature, a low satisfaction level for both needs indicates a similar degree of vulnerability among the people targeted, and the comparison between companies is meaningful. But if the needs are fundamentally different, the people whose needs remain unmet are not in the same situation: someone facing undernutrition and someone lacking access to foreign-language classes do not face the same degree of vulnerability. Without a ranking of the needs being addressed, comparing targeted beneficiaries has no sound basis.

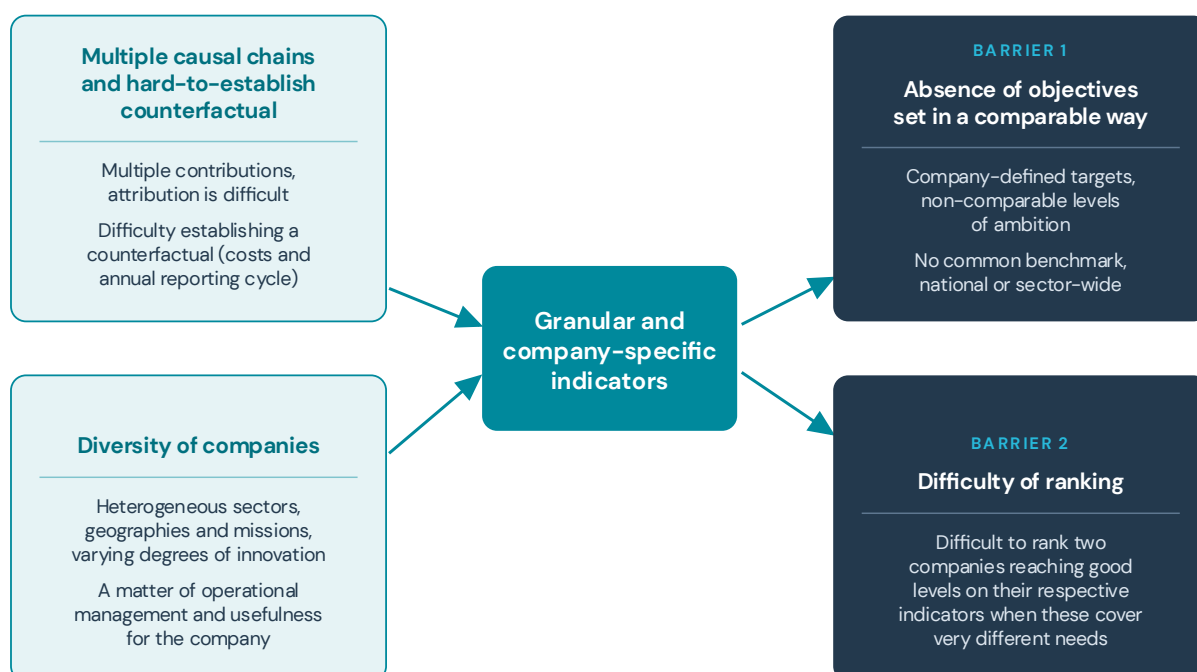


Figure 2. Barriers 1 and 2: The absence of comparable objectives and the difficulty of ranking impacts

2.2 The heterogeneity of data

The literature stresses the need for measurements that are reliable, relevant, and comparable (Arjaliès et al., 2020) — a requirement rarely met in practice. Available approaches are numerous, qualitative or quantitative, experimental or participatory (Leclerc & Sibieude, 2025), and using different methodologies to measure the same object can yield different results.

Field observations

The same object of study — social connection, for example — can be measured through observed interactions, subjective scales administered via questionnaires, or indirect behavioral indicators such as the frequency of participation in group activities. These methods do not measure exactly the same thing, and their results are not directly comparable.

“You can assess how an indicator like self-confidence has changed qualitatively, by asking the person how they perceive their own progress. Some people tell us that’s not valid because it’s subjective. But that perception is exactly what matters. For example, when we measure a person’s isolation (objective), that’s not the same thing as their feeling of loneliness (perception).” — Consulting firm, Impact expert

Data reliability also varies depending on the scale of the studies: some are limited to a few-minute survey, while others draw on several years of observation (Becker, 2001). In addition, the data used to measure social impact are rarely homogeneous. In some cases, investors face a lack of data altogether (Arjaliès et al., 2018). Primary data collected directly from beneficiaries frequently coexist with secondary data extrapolated from existing studies conducted in other contexts, on other populations. Secondary data, however, rest on transferability assumptions that are rarely made explicit. As Pawson & Tilley (1997) point out with the concept of realistic evaluation, the effects of an activity depend on the context in which it takes place and the mechanisms it activates there. Impact is not automatically transferable from one setting to another. As a result, when primary data and secondary estimates are presented side by side without distinction, two companies may display similar results built on very unequal foundations, calling comparability into question.

“When it comes to social impact, we don’t make explicit enough what is actual impact measurement and what is estimation. We put what we track side by side with the estimate.” — Asset manager, Impact expert

”

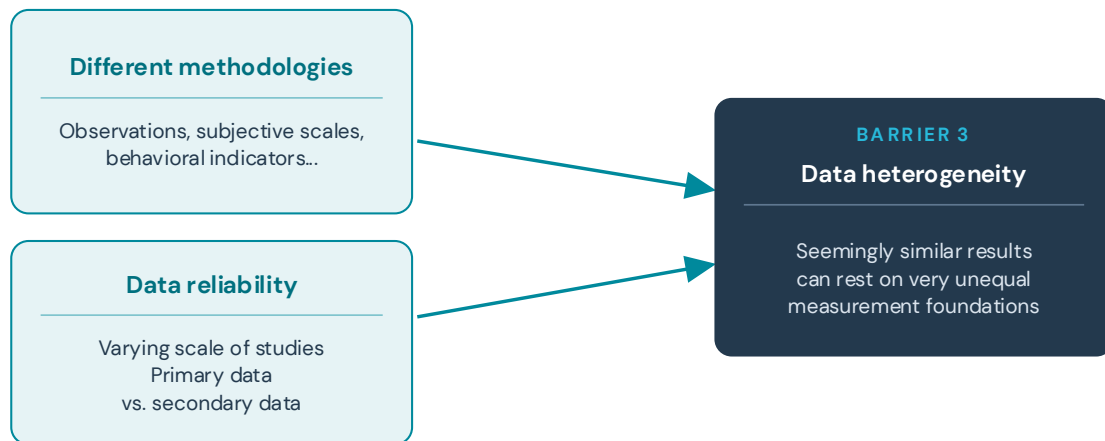


Figure 3. Barrier 3: The heterogeneity of data.

2.3 A further challenge when seeking to aggregate: the question of the fungibility of social impacts

Comparing two distinct impacts becomes even more complicated when one seeks to aggregate them into a single figure representing the overall impact of a company or a portfolio. For certain environmental impacts, fungibility is generally accepted: a tonne of CO₂ sequestered or avoided in one place can be weighed against a tonne of CO₂ emitted elsewhere. On the social side, this raises new questions.

First, between positive and negative social impacts: Mattingly & Berman (2006) show that these two types of impact are empirically and conceptually distinct. The former reflect voluntary, proactive initiatives (e.g., employee training, pollution prevention), while the latter reflect problematic behaviors (e.g., labor disputes, lawsuits, environmental accidents). A company can therefore accumulate both positive and negative behaviors without one offsetting or erasing the other. As a result, aggregating these dimensions would mask factors that do not offset one another, according to the authors.

Field observations

During a working group bringing together impact investing practitioners around approaches to monetizing social impacts, a speaker presented methodologies for expressing impacts in monetary terms: on one hand, a euro value of remediation, answering the question “how much would it cost to repair the harm caused to the affected parties?”; on the other, a euro value of well-being, answering the question “to achieve the same improvement in felt well-being, how much additional income would need to have been given to them?”

At the end of the presentation, the speaker raised the question of whether these two types of value could be compared and offset one another. The answer was no: “They are

in euros but different euros” — although expressed in the same unit, these values capture fundamentally different realities, one negative (remediation of harm) and the other positive (improvement in well-being), affecting, in his view, distinct groups of people. They cannot therefore be aggregated into a net benefit without masking what sets them apart.

This position sparked discussion in the room, with some practitioners pointing out that aggregation was precisely the purpose of monetizing impacts.

In an entirely different context, an impact expert also raised questions about offsetting between social impacts:

“For a company that does at-home appliance repairs, the point is to check that the people traveling to make the repair won’t create a negative impact greater than the positive impact of the repair itself. We then look at the ratio of tCO2 relative to a new machine. On social topics that works less well: saving lives to offset other deaths — that doesn’t work.” — Asset manager, Impact expert

The question also arises between two positive impacts. Accepting that improvement in an older person’s physical well-being and improvement in their social connection can be added together assumes that a gain on one can offset a shortfall on the other. Similarly, aggregating into a single score the impact of one company on women’s health and that of another on the professional integration of at-risk youth requires establishing a compensatory relationship between two fundamentally different realities.

This indeterminacy, moreover, is not limited to impacts of a different nature or opposite sign; it is present even in the simple addition of occurrences of the same indicator. Adding up the number of people trained, or beneficiaries supported, amounts to treating each person as an interchangeable unit, equivalent to all others. Yet even with comparable profiles and in the same context, each individual situation reflects a singular reality that the total erases.

This question of fungibility is neither neutral nor purely technical. It rests on unresolved normative debates that stand in the way of any consensual aggregation of social impacts.

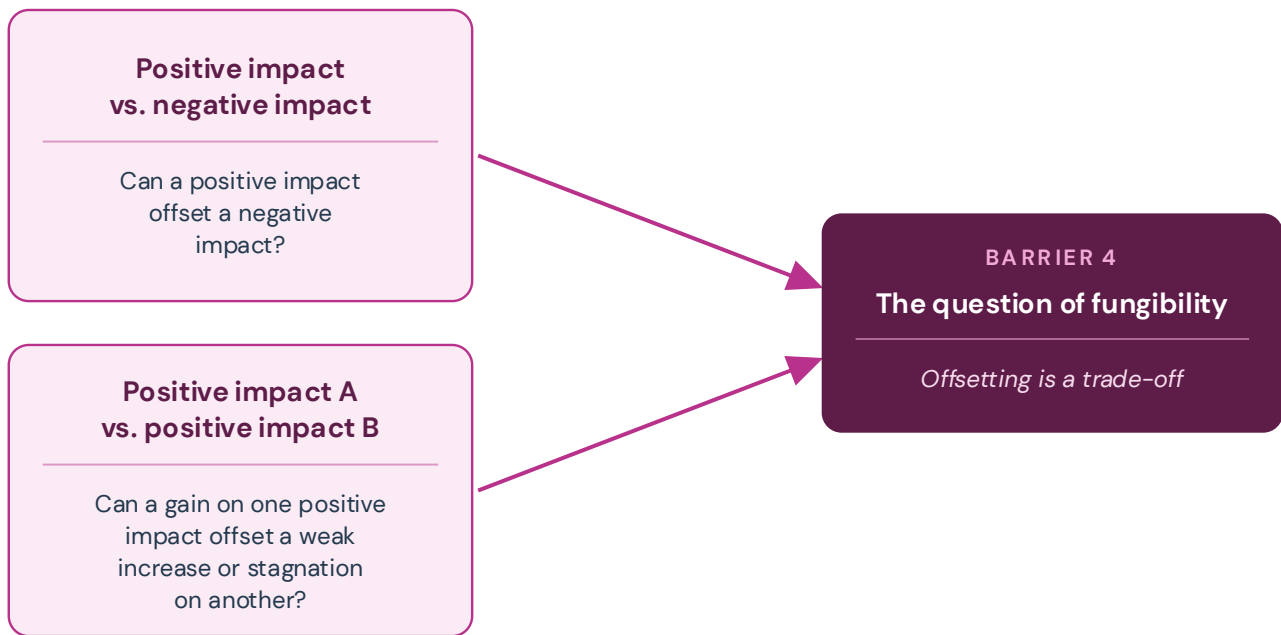


Figure 4. Barrier 4: The question of the fungibility of social impacts.

The diagram below summarizes how these four barriers fit together:

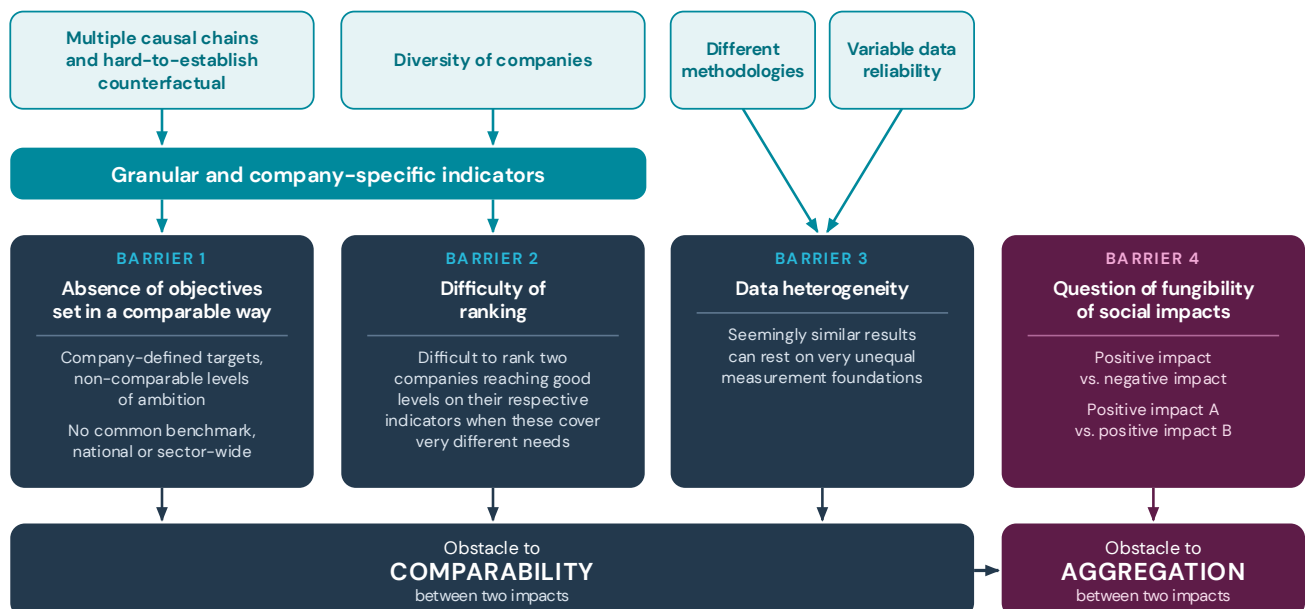


Figure 5. Summary: The barriers to comparing and aggregating two social impacts

3. Comparing two companies: an additional obstacle

The difficulties set out above at the level of comparing two distinct impacts accumulate when one seeks to compare two companies with one another. A further obstacle specific to this level of analysis is added: the heterogeneity of measurement scopes.

3.1 An additional obstacle: the heterogeneity of measurement scopes

Impact assessments are subject to computational complexity because a large number of elements can potentially be taken into account (Stievenart & Pache, 2014). An impact measurement process involves prioritizing certain pieces of information over others, so the results are necessarily incomplete. This inevitably introduces a degree of bias.

Field observations

A company operating in the personal-care sector affects many stakeholders: direct beneficiaries, of course; but also their relatives and caregivers, whose burden the intervention may ease; partner healthcare professionals, whom the company supports or complements in their work; local communities, if the company has a strong local presence; and public authorities, concerned with the costs avoided that the intervention may generate.

Since measurement exercises are resource- and time-intensive, it is not possible to cover every stakeholder within such a process. A scope is therefore prioritized. This scope can be defined according to several variable factors: the company's intentionality (which effects does it deliberately seek to generate, at the core of its strategy?), the perceived materiality of the issues at stake (which stakeholders are most affected?), and communication choices. It may also evolve over time, expanding as the company gains maturity in its measurement practice.

"We chose to target beneficiaries. There are lots of stakeholders, but from the very beginning we made the choice to collect impact data for this particular stakeholder group, because we really wanted to understand and showcase the added value of our content, formats, speakers, teaching methods, and so on." — Impact expert, Company

"We haven't launched any measurement on activity coordinators because we can't do everything at once, but of course our services offer them a broader portfolio of activities." — Impact expert, Company

This barrier, moreover, is not limited to the heterogeneity of indicators — it also arises between homogeneous indicators. Two companies reporting the same indicator — the number of jobs supported, for example — are not directly comparable if one counts only its direct employees while the other also includes the indirect jobs in its value chain.

This problem is analogous, in the climate domain, to the distinction between scope 1, 2, and 3 greenhouse gas emissions. A company that measures only its direct emissions is not directly comparable to one that also includes indirect emissions linked to its supply chain or the downstream use of the products it sells: the scope chosen fundamentally changes what is being measured.

Without a shared framing of the measurement scope, the gaps observed between companies may thus reflect genuine differences in impact, or simply different scope choices.

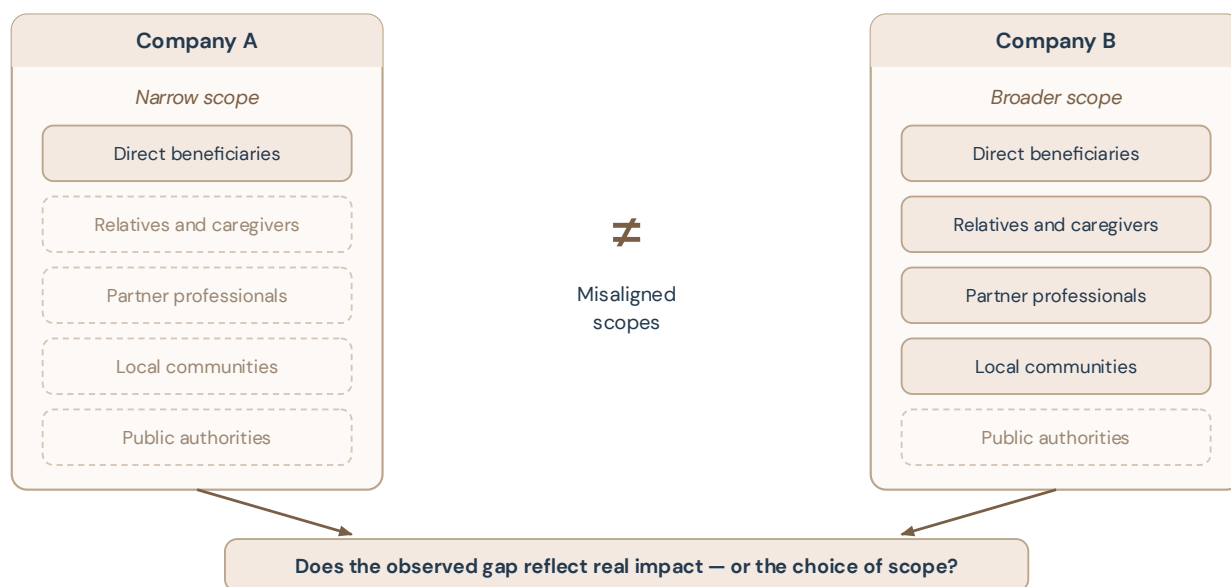


Figure 6. Barrier 5: The heterogeneity of measurement scopes

3.2 Inter-company comparison: an accumulation of every obstacle

Inter-company comparison accumulates all the obstacles identified at the level of individual impacts — indicators that are granular and company-specific, which prevent access to shared thresholds and complicate ranking; heterogeneity of methods and data — and adds to them the obstacle of scope. Even two companies operating in the same sector with similar impact intentions can be difficult to compare directly if their indicators differ, their collection methods diverge, and their scopes are not delineated according to the same principles.

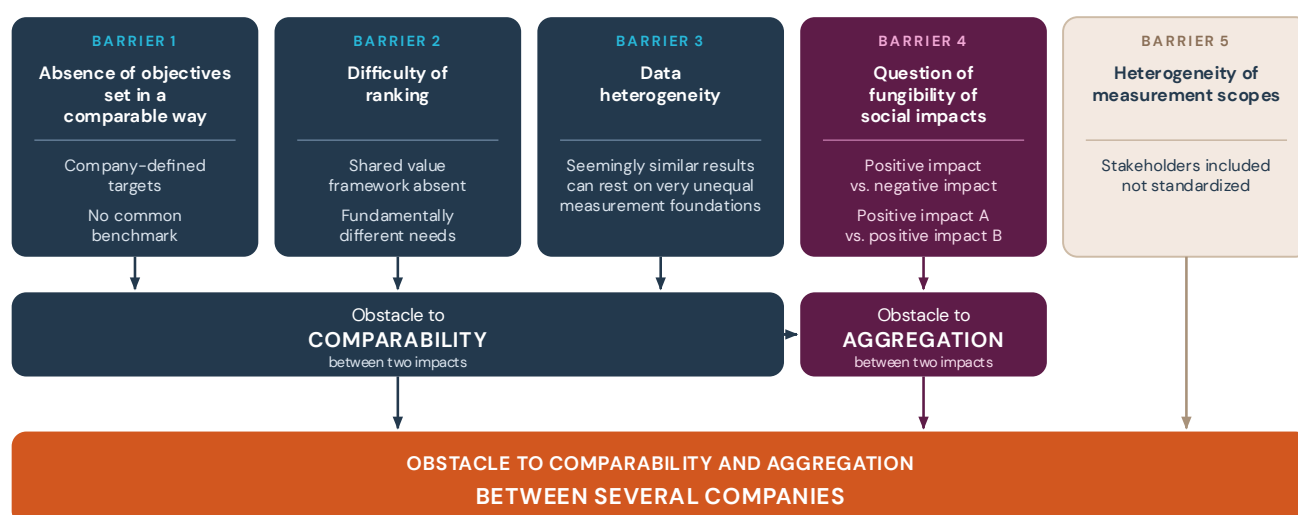


Figure 7. Summary: The barriers to comparing and aggregating social impacts across multiple companies

For the impact investor, who must nonetheless make allocation decisions across these companies, this situation is demanding. It calls for approaches that seek to build frameworks capable of positioning impacts relative to one another, so as to avoid losing granularity through standardization.

4. Conclusion: avenues toward comparable social impact measurement

This white paper has explored the structural barriers that make comparing and aggregating heterogeneous indicators difficult, through the case of social impact. The tension between the specificity of indicators and the need for consolidation lies at the heart of the problem. Consolidation practices today rely mainly on volume indicators — number of beneficiaries, jobs created — that apply across a wide range of activities. But these indicators do not capture the quality of the impact or the vulnerability of the targeted populations. The challenge, then, is to overcome these barriers, which calls for several complementary avenues of work. It should be stressed from the outset, however, that the scale of these undertakings depends closely on the level at which one seeks to compare: comparing within a fund, between funds of the same manager, or between managers each require different degrees of coordination.

Ranking targeted impacts and beneficiary vulnerability. Faced with fundamentally heterogeneous indicators, a first avenue consists of building dimensions for ranking impacts according to shared principles: to what extent is a given impact considered more fundamental than another? What is the vulnerability of the targeted populations? These choices involve normative trade-offs that must be made explicit and openly owned.

Some investors are already moving in this direction, seeking to characterize the socio-economic vulnerability of their beneficiaries. For example, the percentage of beneficiaries covered by a public assistance scheme serves as a proxy for their degree of vulnerability — not a direct ranking indicator, but an attempt to qualify the intensity of the need being addressed.

“For the characterization, we ask how long the person has been in the facility, whether they are widowed, divorced, or living alone, in order to capture their family situation. We clearly identify where the workshop takes place, whether it’s in an underserved area with very few activities available. We’re also working on identifying a bit better the difficulties these beneficiaries face, and how closely we’re able to respond to that specific audience. For now, we have this sense and perception that we’re aware of, but we haven’t really pinned it down yet by saying, ‘this is really what will let us distinguish between one action and another.’” —
Impact expert, Company

”

This avenue nonetheless remains difficult to operationalize. Vulnerability is multidimensional: identifying which dimension takes precedence over the others in order to establish a ranking also requires normative choices. Similarly, ranking impacts against one another — according to the needs they address — may seem obvious in some cases, but can become highly delicate in others (emergency care versus mental health, health versus education). These normative questions have no technical answer.

Defining objectives and thresholds according to shared principles. A second avenue consists of harmonizing how impact objectives are set and measured. When the quality of an impact is measured using Likert scales¹ administered to beneficiaries, these scales are by definition standardized and bounded — thereby offering a well-defined basis for comparison, independent of the indicator measured. This approach is one example of a promising avenue for building comparable impact indicators across companies, without requiring standardization of the indicators themselves.

Harmonizing data-collection methodologies. A gradual convergence of practices — in survey protocols, the scales used, the distinction between primary and secondary data — would help strengthen the comparability of measurements. This undertaking requires collective work whose scale varies depending on the intended level: within a fund, it involves alignment between the investor and its companies; between funds or between managers, it requires much broader market-wide coordination.

Standardizing or making scopes explicit. Consistently defining and justifying which stakeholders are included in the measurement scope — and to what extent these scopes can be considered comparable with one another — helps ensure that comparisons rest on a homogeneous basis, rather than on unexplained framing choices.

Finally, the limits of this exercise must be acknowledged. This white paper is, above all, a diagnosis: it explains why comparing and aggregating heterogeneous impacts is difficult, and why these barriers are not easily resolved — social impact represents an extreme case, one that combines the full range of these difficulties along with a degree of heterogeneity found, to a lesser extent, in other domains. The avenues outlined above mitigate these barriers without making them disappear: a fully rigorous aggregation of social impacts almost certainly does not

¹ A psychometric measurement scale made up of ordered response levels (for example, from “strongly disagree” to “strongly agree”), used to quantify the degree of agreement, satisfaction, or perception of a respondent regarding a given statement.

exist. That said, the subject does not end with this observation. In practice, it falls to each investor to define a pragmatic — necessarily imperfect — approach that nonetheless allows them to report on their impacts, guide their decisions, and steer their portfolios. The real question, then, is not so much whether these approaches are perfect, but rather what imperfections and biases are acceptable given the exercise being conducted and the use that will be made of it. Social impact measurement always involves choices — on scope; on indicators and their potential aggregation, and the weightings applied; on data collection methods — all of which influence the results obtained. Making these choices explicit, and clearly distinguishing what stems from direct measurement from what stems from estimation, is a condition for the credibility and usefulness of social impact measurement.

These questions form part of a broader research program conducted within the i3-CRG laboratory, École polytechnique, CNRS, Institut polytechnique de Paris, Palaiseau.

References

- Arjaliès, D.-L., Bouchet, V., Crifo, P., & Mottis, N. (2020). La mesure d'impact et l'investissement socialement responsable (ISR): Un tour d'horizon [Impact measurement and socially responsible investment (SRI): An overview]. In E. Tchotourian, L. Bres, & L. Geelhand de Merxem (Eds.), *Zone frontières et entreprise socialement responsable—Perspective multiple: Droit, administration et éthique*. Yvon Blais; Mare & Martin. <https://papers.ssrn.com/abstract=3733755>
- Arjaliès, D.-L., Chollet, P., Crifo, P., & Mottis, N. (2018). *Mesure d'impact et label ISR: Analyse et recommandations* [Impact measurement and the SRI label: Analysis and recommendations]. Comité Scientifique du Label ISR français. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3559718
- Becker, H. A. (2001). Social impact assessment. *European Journal of Operational Research*, 128(2), 311–321. [https://doi.org/10.1016/S0377-2217\(00\)00074-6](https://doi.org/10.1016/S0377-2217(00)00074-6)
- Comité Français sur l'investissement à impact social. (2014). *Comment et pourquoi favoriser des investissements à impact social? Innover financièrement pour innover socialement* [How and why to promote social impact investments? Innovating financially to innovate socially]. https://www.economie.gouv.fr/files/files/PDF/RapportSIIFce_vdef_28082014.pdf
- Corvo, L., Pastore, L., Manti, A., & Iannaci, D. (2021). Mapping social impact assessment models: A literature overview for a future research agenda. *Sustainability*, 13(9), Article 4750. <https://doi.org/10.3390/su13094750>
- Crucke, S., & Decramer, A. (2016). The development of a measurement instrument for the organizational performance of social enterprises. *Sustainability*, 8(2), Article 161. <https://doi.org/10.3390/su8020161>
- Ebrahim, A., & Rangan, V. K. (2014). What impact? A framework for measuring the scale and scope of social performance. *California Management Review*, 56(3), 118–141. <https://doi.org/10.1525/cmr.2014.56.3.118>
- Ferreiro, M., & Kölbel, J. (2025). *Impact measurement harmonization: Challenges and opportunities*. Universität St. Gallen. https://fsi.unisg.ch/fileadmin/user_upload/HSG_ROOT/Institut_FSI/Dokumente/Sustainable_Finance/Final_Impact_Measurement_Harmonization_report.pdf
- France Invest & FIR. (2021). *Investissement à impact: Une définition exigeante pour le coté et le non-coté* [Impact investing: A demanding definition for listed and unlisted assets]. <https://www.franceinvest.eu/definition-exigeante-de-linvestissement-a-impact-pour-le-cote-et-le-non-cote/>
- Höchstädter, A. K., & Scheck, B. (2015). What's in a name: An analysis of impact investing understandings by academics and practitioners. *Journal of Business Ethics*, 132(2), 449–475. <https://doi.org/10.1007/s10551-014-2327-0>
- Kaufmann, L., Krlev, G., & Brown, M. M. (2025). In plain sight: Mechanisms of means–ends decoupling in impact investing. *Organization Studies*, 46(5), 667–692. <https://doi.org/10.1177/01708406241295505>
- Kroeger, A., & Weber, C. (2014). Developing a conceptual framework for comparing social value creation. *Academy of Management Review*, 39(4), 513–540. <https://doi.org/10.5465/amr.2012.0344>

- Leclerc, É., & Sibieude, T. (2025). *La mesure d'impact: Intégrer l'impact social et environnemental dans sa performance économique* [Impact measurement: Integrating social and environmental impact into economic performance] (1st ed.). Dunod.
- Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, 45(1), 20–46. <https://doi.org/10.1177/0007650305281939>
- Mulgan, G. (2010). Measuring social value. *Stanford Social Innovation Review*, 8(3), 38–43. <https://doi.org/10.48558/NQTO-DD24>
- Nicholls, A. (2009). 'We do good things, don't we?': 'Blended value accounting' in social entrepreneurship. *Accounting, Organizations and Society*, 34(6), 755–769. <https://doi.org/10.1016/j.aos.2009.04.008>
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. SAGE Publications.
- Sibieude, T., Leclerc, É., Leclair, C., Neyron, C., Raynaud de Lage, L., de Rochechouart, L., Dupuis, E., Gros, C., & Kleynjans, O. (2021). *Petit précis de l'évaluation de l'impact social #2* [A short guide to social impact evaluation #2]. ESSEC Business School, Avise, Mouvement Impact France. https://www.avise.org/sites/default/files/atoms/files/20211203/1223361021_petit_precis_de_levaluation_de_limpact_social_2_web_1.pdf
- Stievenart, E., & Pache, A.-C. (2014). Évaluer l'impact social d'une entreprise sociale: Points de repère [Evaluating the social impact of a social enterprise: Benchmarks]. *Revue internationale de l'économie sociale: Recma*, (331), 76–92. <https://doi.org/10.7202/1023486ar>
- World Economic Forum. (2013). *From the margins to the mainstream: Assessment of the impact investment sector and opportunities to engage mainstream investors*. <https://www.weforum.org/publications/margins-mainstream-assessment-impact-investment-sector-and-opportunities-engage-mainstream-investors/>