

Les freins à l'agrégation d'indicateurs d'impact hétérogènes

Le cas de l'impact social

White Paper

Auteur : Axelle Guers

Résumé

La mesure de l'impact dans l'investissement à impact reste un champ méthodologique en construction. Sur certains sujets environnementaux, comme le climat, des métriques établies et partagées existent — les tonnes de CO₂ induites, évitées ou séquestrées bénéficient par exemple aujourd'hui de standards et de normes comptables qui en facilitent la comparaison et la consolidation. L'impact social, lui, illustre de façon saillante une difficulté plus générale : celle de comparer et de consolider des indicateurs hétérogènes. Ses indicateurs sont souvent granulaires, spécifiques aux activités de chaque entreprise, et difficilement comparables entre eux. En pratique, les investisseurs parviennent à consolider à l'échelle d'un fonds des indicateurs de volume — nombre de bénéficiaires touchés, emplois soutenus —, relativement homogènes ; mais les indicateurs qui renseignent sur les changements générés sur les parties prenantes demeurent, eux, propres à chaque entreprise et difficilement agrégeables. C'est précisément sur ces derniers que la comparaison est la plus difficile, alors qu'ils constituent l'information la plus riche sur ce que produit l'intervention de l'entreprise.

Ce document explore pourquoi la comparaison et la consolidation d'indicateurs hétérogènes constituent un défi structurel pour les investisseurs à impact, et identifie les mécanismes qui en sont à l'origine. Pour étudier ce défi, nous retenons le cas des indicateurs d'impact social, particulièrement révélateurs de cette difficulté.

Nous identifions cinq freins principaux. Tout d'abord, le caractère granulaire et spécifique des indicateurs sociaux donne lieu à (1) une absence d'objectifs d'impact établis selon des principes comparables — comparer le taux d'atteinte d'un objectif peu ambitieux à celui d'un objectif plus exigeant fixé selon des principes différents est délicat — et à (2) une difficulté à hiérarchiser des impacts hétérogènes entre eux — une amélioration de la santé vaut-elle une amélioration du niveau en langues vivantes ? Par ailleurs, (3) l'hétérogénéité des données — méthodologies et degrés de fiabilité des données variables — pose une question de comparabilité des résultats de mesure : deux valeurs en apparence similaires peuvent reposer sur des bases très inégales. Se pose ensuite la question de (4) la fongibilité des impacts sociaux : agréger des impacts et établir leur substituabilité suppose des arbitrages normatifs qui ne peuvent être résolus uniquement sur le plan technique. Enfin, lorsque l'on cherche à comparer deux entreprises, (5) l'hétérogénéité des périmètres de mesure fait que les écarts observés peuvent refléter des choix de cadrage autant que des différences d'impact réel. Ces trois derniers freins valent aussi pour des indicateurs ayant une même unité, mais c'est leur accumulation sur des indicateurs hétérogènes qui fait la difficulté centrale.

En conclusion, nous établissons plusieurs pistes de réflexion pour lever ces freins : hiérarchiser les impacts et la vulnérabilité des bénéficiaires, définir des objectifs selon des principes communs, harmoniser les méthodologies de collecte, et expliciter les périmètres de mesure. Ces pistes appellent toutes à une plus grande transparence sur les hypothèses sous-jacentes aux mesures d'impact social.

Introduction

L'investissement à impact désigne une approche d'investissement qui cherche délibérément à générer à la fois un rendement financier et un impact social ou environnemental positif et mesurable (World Economic Forum, 2013). Il se distingue de l'investissement socialement responsable, qui vise principalement à réduire les risques extra-financiers (Höchstädter & Scheck, 2015), par trois piliers spécifiques : l'intentionnalité ex ante de l'investisseur, l'additionnalité de son action, et la mesure des externalités sociales et environnementales générées (France Invest & FIR, 2021).

Les impacts visés sont variés, et ce d'autant plus dans le domaine social, sur lequel porte ce white paper : création d'emplois dans des zones défavorisées, amélioration de l'accès à l'éducation, à la santé ou à l'énergie pour des populations mal desservies (France Invest & FIR, 2021). Cette diversité est constitutive du champ et se trouve au niveau des entreprises financées : un même fonds peut simultanément financer et accompagner une entreprise proposant des soins de santé pluridisciplinaires pour les femmes dans des zones où l'offre médicale spécialisée est insuffisante, une plateforme numérique d'orientation et de développement des compétences accessible à des bénéficiaires à travers le monde, notamment dans des pays du Sud, et une entreprise organisant des ateliers de prévention et d'animation pour lutter contre l'isolement des personnes âgées. Chacune mesure son impact avec des indicateurs qui lui sont propres — délais d'accès aux soins réduits et satisfaction des patientes pour la première, progression de la confiance en soi dans le parcours professionnel pour la deuxième, proportion de bénéficiaires déclarant se sentir moins isolés pour la troisième — et ces indicateurs sont fondamentalement différents les uns des autres.

Ce white paper se positionne précisément sur ce cas de figure : la comparaison et la consolidation d'indicateurs hétérogènes, construits pour des activités et des missions fondamentalement différentes. La difficulté soulevée par cette hétérogénéité n'est pas propre au social, elle se pose à tout investisseur cherchant à rendre compte de son impact de façon agrégée, en particulier lorsqu'il gère un portefeuille multi-sectoriel. L'impact social constitue toutefois un terrain d'étude privilégié, les effets générés sur les parties prenantes étant particulièrement divers.

Définition

Il convient ici de préciser que nous définissons, dans le cadre de ce white paper, l'impact social comme les effets générés par une intervention sur ses parties prenantes.

La littérature distingue trois niveaux d'indicateurs :

- Les indicateurs d'output (de réalisation) : les produits ou services directement délivrés par l'activité de l'entreprise – biens distribués, personnes formées, bénéficiaires touchés, etc.

- Les indicateurs d'outcome (de résultat) : les changements à moyen et long terme que connaissent les bénéficiaires à la suite de l'intervention.
- Les indicateurs d'impact : les changements significatifs et durables sur les causes profondes, ou à l'échelle des communautés ou de la société (Ebrahim & Rangan, 2014).
- Dans ce document, nous nous concentrons principalement sur les indicateurs d'outcome et d'impact, que nous désignons de façon commune par le terme « indicateur d'impact social ». Ce terme inclut une dimension de qualité et de nature de l'effet produit, et non uniquement son volume, à la différence de la mesure d'output.

Comme le soulignent Kaufmann et al. (2025), la mesure d'impact est progressivement devenue une caractéristique essentielle des acteurs du champ, allant jusqu'à faire partie des pratiques désormais attendues pour opérer et être identifié dans l'écosystème de l'investissement à impact — et ce, avant même que des méthodologies cohérentes n'aient été stabilisées dans le domaine social. La littérature sur l'évaluation d'impact, bien qu'ancienne — ses origines remontent à l'après-Seconde Guerre mondiale, dans le contexte de l'essor des politiques publiques sociales puis de la structuration de l'aide au développement (Sibieude et al., 2021) — reste fragmentée et constituée de méthodologies hétérogènes qui n'ont pas été conçues pour les enjeux spécifiques de l'investissement à impact (Corvo et al., 2021).

Cette situation crée une tension particulière pour les investisseurs, confrontés à deux exigences simultanées : piloter l'impact des entreprises en portefeuille au plus près de leur mission opérationnelle, et rendre compte de cet impact de façon synthétique à leurs parties prenantes (Arjaliès et al., 2018).

Faute de standards communs, les objectifs de transparence, de comparabilité ou encore d'amélioration de la performance sociale restent ainsi difficiles à atteindre. Ce constat appelle à interroger les fondements à l'origine de cette désarticulation. Ce white paper explore les freins structurels à la comparaison et à la consolidation des impacts sociaux.

Méthodologie

Ce white paper s'appuie sur la littérature académique sur la mesure d'impact social et l'investissement à impact et sur des observations de terrain conduites au sein de Mirova, sur une période de seize mois, complétées par dix-neuf entretiens semi-directifs avec des professionnels de l'investissement à impact au sein de différentes structures — gestionnaires d'actifs, investisseurs institutionnels, entreprises et cabinets de conseil. Les verbatims et exemples cités dans ce document sont issus de ces entretiens et observations de terrain, et visent à illustrer et compléter certains éléments issus de la littérature.

1. Dans l'investissement à impact : un besoin de comparer et de consolider les impacts

Les sociétés de gestion à impact sont confrontées à trois objectifs distincts (Arjaliès et al., 2018).

Le premier est le **pilotage de l'impact** des entreprises accompagnées. L'investisseur suit l'évolution de chaque entreprise au regard de ses objectifs d'impact, ce qui suppose de disposer d'indicateurs lisibles et adaptés à ses activités.

Le deuxième est le **reporting** aux investisseurs. Les sociétés de gestion jouent un rôle d'intermédiaire financier et doivent synthétiser l'information relative à l'impact pour rendre des comptes. Cette contrainte appelle des indicateurs non seulement consolidés et communicables à l'échelle d'un portefeuille, mais aussi suffisamment lisibles et comparables d'une société de gestion à l'autre (Corvo et al., 2021).

Sur un portefeuille social on va avoir des indicateurs individuels mais qu'est-ce qu'on va raconter à l'échelle du portefeuille ? L'idée c'est d'avoir quelques grands chiffres. — Asset manager, Expert de l'impact

”

Le troisième est l'**arbitrage d'allocation** : les investisseurs ont également besoin de comparer des entreprises entre elles afin d'effectuer des arbitrages à l'échelle d'un fonds tout en tenant compte de l'impact.

Tu as 1€ à investir, tu investis où ? La question qu'on doit se poser c'est celle de la maximisation de notre impact en tant que financeur. Comment classer ces entreprises par rapport à ce qu'elles créent, par euro investi ? — Asset manager, Expert de l'impact

”

En pratique, la consolidation d'indicateurs sociaux à l'échelle des fonds repose aujourd'hui principalement sur des indicateurs de volume — nombre de bénéficiaires touchés, emplois soutenus — car ils sont transversaux à des activités très variées. Mais le volume seul ne rend pas compte de la qualité de l'impact ni de l'ensemble de la valeur sociale créée : un accès à l'eau potable pour des ménages ruraux sans infrastructure de base n'est pas directement comparable à des cours de cuisine végétarienne pour des urbains sensibilisés aux enjeux alimentaires. De même, un programme d'alphabétisation pour des adultes en situation d'illettrisme ne se compare pas directement à une application d'apprentissage de langues vivantes pour lycéens. Le volume seul ne permet pas de rendre compte de ces différences.

C'est précisément cette rencontre entre la spécificité nécessaire des indicateurs au niveau des entreprises et le besoin de synthèse au niveau du fonds qui crée la tension centrale explorée dans ce white paper.

2. Comparer deux impacts : les freins à l'œuvre

2.1. L'absence d'objectifs comparables et la difficulté de hiérarchisation

Cette section met en évidence deux freins qui découlent tous deux du caractère granulaire et spécifique des indicateurs sociaux : **l'absence d'objectifs d'impact établis de façon comparable**, et **la difficulté à hiérarchiser des impacts hétérogènes entre eux**. Pour comprendre l'origine de ces deux freins, il faut d'abord revenir sur les conditions qui permettraient, en théorie, de comparer la valeur sociale créée par différentes entreprises.

Idéalement, pour comparer la valeur sociale créée par différentes entreprises, on souhaiterait disposer de trois éléments (Kroeger & Weber, 2014) :

Des dimensions d'impact comparables en importance. Pour que la comparaison entre deux interventions soit valide, il faut qu'elles soient évaluées sur des dimensions qui ont un poids équivalent pour le bien-être des bénéficiaires. C'est le principe que posent Kroeger & Weber (2014) en proposant de mesurer la satisfaction des bénéficiaires dans différents domaines de leur vie — santé, éducation, lien social — et de s'assurer que les domaines retenus sont suffisamment proches en importance.

Des objectifs définis de façon comparable. Pour que la comparaison entre deux interventions soit pertinente, il faut pouvoir situer leurs résultats par rapport à un point de référence commun — un seuil définissant ce qui constitue une situation de besoin, et par rapport auquel on peut mesurer si et dans quelle proportion l'écart a été réduit. Sans ce seuil partagé, on peut décrire ce qu'une intervention produit en valeur absolue, mais pas évaluer son efficacité ni la comparer à celle d'une autre intervention opérant dans un contexte différent.

Kroeger & Weber (2014) proposent à cet effet d'utiliser la moyenne nationale de satisfaction dans un domaine de vie — santé, éducation, lien social — comme référence : les individus en dessous de cette moyenne sont considérés comme en situation de besoin, et le degré de valeur sociale créée se mesure par la réduction de cet écart. Cette approche permet notamment de comparer l'efficacité de deux interventions en rapportant leurs résultats à des objectifs établis de façon cohérente avec leur contexte : si une population est plus difficile à atteindre, le seuil de référence le reflétera naturellement, rendant la comparaison plus pertinente. Kroeger & Weber proposent cette logique à l'échelle nationale, mais elle pourrait en principe être déclinée à une granularité plus fine — une région, une population spécifique.

Cette logique est déjà à l'œuvre dans certaines pratiques de l'investissement à impact. Sur des indicateurs comme le revenu ou le salaire décent, les investisseurs utilisent couramment des benchmarks géographiques, par exemple ceux produits par l'Anker Research Institute. Le niveau cible varie selon la géographie, mais c'est précisément ce qui rend la comparaison pertinente : on ne compare pas des niveaux absolus, mais des écarts à un seuil ancré dans les réalités locales. L'enjeu, plus largement, est de disposer d'objectifs définis selon une logique commune, permettant de comparer des niveaux d'atteinte entre interventions opérant dans des environnements différents.

Un contrefactuel permettant d'isoler l'effet propre de l'intervention de ce qui se serait passé sans elle — sans lui, il est difficile de distinguer ce qui est dû à l'action de l'entreprise de ce qui

résulte de facteurs externes comme la conjoncture, les politiques publiques en cours, ou les dynamiques propres à la population ciblée.

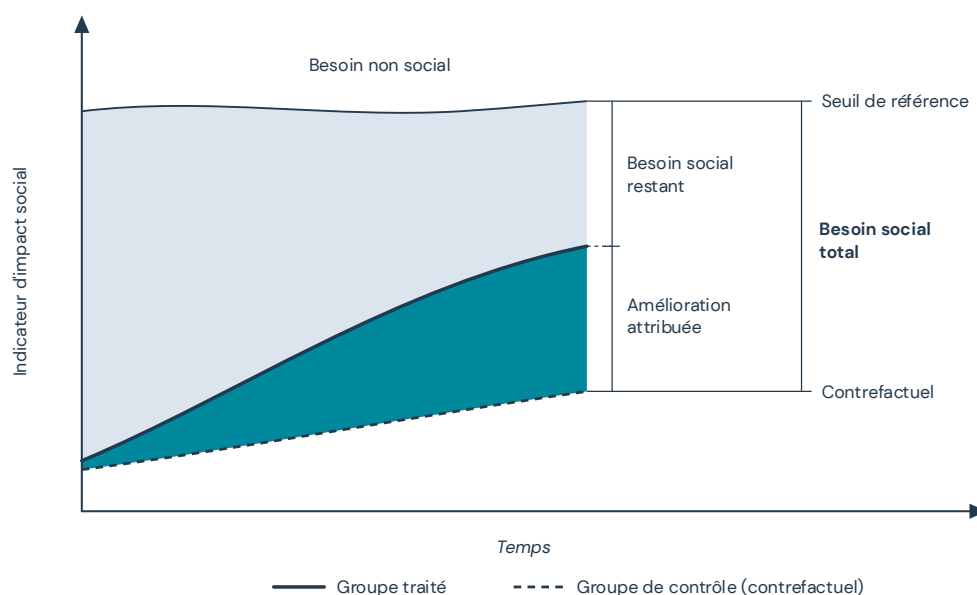


Figure 1. Cadre pour comparer la valeur sociale créée (adapté de Kroeger & Weber, 2014).

Dans ce cadre, le degré de valeur créée correspond à l'amélioration attribuée divisée par le besoin social total.

Ces trois conditions sont rarement réunies dans le contexte de l'investissement à impact.

2.1.1 Les chaînes causales multiples et la difficulté du contrefactuel

Avant de formaliser les deux freins qui en résultent (section 2.1.3), il convient de revenir sur leur origine : la difficulté du contrefactuel et le recours à des indicateurs spécifiques aux activités des entreprises.

Dans le domaine social, les résultats observés résultent rarement de l'action d'un seul acteur : plusieurs organisations, politiques publiques et dynamiques contextuelles contribuent simultanément à un même résultat, rendant très difficile la distinction de la contribution propre d'une entreprise donnée. C'est précisément le rôle du contrefactuel : en comparant le groupe traité à un groupe de contrôle aux caractéristiques similaires, il isole l'effet propre de l'intervention des facteurs externes.

Observations de terrain

Le contrefactuel idéal est structurellement difficile à construire dans le contexte de l'investissement à impact. La collecte de données auprès d'un groupe de contrôle requiert des ressources importantes, s'ajoutant à la collecte sur les

bénéficiaires directs, et sur l'ensemble des entreprises d'un portefeuille. Les rythmes de reporting annuels laissent peu de place à des collectes lourdes et fréquentes. En pratique, les investisseurs recourent à des approximations comme le contrefactuel auto-déclaratif, qui renseigne sur ce que le bénéficiaire attribue à l'intervention.

« Ceux qui ne connaissent pas se disent « Pourquoi on ne fait pas un groupe de comparaison ? ». On est sur un terrain social, multifactoriel, monter un contrefactuel est une méthode qui est lourde, onéreuse et pas adaptée à l'environnement. Une méthode un peu moins robuste est le « avant-après », mais là aussi c'est une méthode qui n'est pas toujours adaptée par rapport aux ressources disponibles et par manque d'anticipation. Nous privilégions des approches pragmatiques pour aider la structure à répondre aux questions évaluatives qu'elle se pose, comprendre et mesurer son impact en essayant d'apporter une rigueur tout au long du process pour éviter les biais. Donc nous pratiquons beaucoup des méthodes mixtes qualitatives et quantitatives souvent en ex post. Bien sûr on distingue le ressenti à chaud, qui comporte des biais et le ressenti après 3 ou 6 mois. On pose alors des questions qui permettent d'attribuer les changements : « Avez-vous le sentiment d'avoir développé telles compétences ? Est-ce grâce au programme ? » » - Cabinet de conseil, Expert de l'impact

Cette approche a cependant une limite importante : la pertinence de la réponse dépend étroitement de la précision de la question posée. Si l'on demande à un participant à une semaine de formation professionnelle dans quelle mesure son niveau d'éducation général s'est amélioré, la question est trop large pour être pertinente — l'individu peut rencontrer des difficultés à isoler l'effet de cette formation de l'ensemble de son parcours. En revanche, si la question porte précisément sur les compétences visées par la formation — par exemple, sa capacité à rédiger un CV ou à conduire un entretien — la réponse devient plus fiable et davantage attribuable à l'intervention. La qualité du contrefactuel auto-déclaratif repose donc sur la granularité et la précision des questions administrées.

Ebrahim & Rangan (2014) formalisent cette difficulté et montrent, à partir de deux cas qu'ils ont étudiés, que l'attribution causale n'est crédible que sous deux conditions. Soit le lien entre l'action de l'organisation et le résultat est très bien documenté scientifiquement : c'est le cas de l'Aravind Eye Hospital en Inde, dont la chirurgie de la cataracte restaure directement et de façon mesurable la vision des patients. Soit l'organisation maîtrise elle-même l'ensemble du parcours d'intervention nécessaire pour produire le résultat : c'est le cas du Harlem Children's Zone à New York, qui accompagne les enfants d'un quartier défavorisé de la naissance jusqu'à l'entrée dans l'emploi, en intégrant éducation, soutien familial, santé et insertion professionnelle dans un périmètre géographique délimité.

Ces deux conditions sont exceptionnelles. Pour la grande majorité des entreprises à impact, les résultats finaux résultent de contributions multiples qu'aucun acteur seul ne maîtrise.

Observations de terrain – L'investisseur à impact : un agrégateur qui ne maîtrise pas la chaîne

Ebrahim & Rangan suggèrent que ce sont les financeurs — fondations, gouvernements, agences de développement — qui sont les mieux positionnés pour mesurer les impacts systémiques, car ils ont une vue d'ensemble sur plusieurs organisations contribuant à un même objectif.

Cette logique ne se transpose pas directement aux investisseurs à impact. Certains fonds se structurent autour d'une thématique commune — la santé, l'insertion professionnelle, l'accès à l'énergie — ce qui crée une cohérence d'ensemble. Pour autant, les entreprises interviennent souvent dans des géographies variées, auprès de populations distinctes, et ne s'inscrivent pas nécessairement dans un parcours d'intervention coordonné sur un territoire délimité. Certains fonds adoptent par ailleurs une approche multi-sectorielle, ce qui les éloigne encore davantage de la logique d'écosystème intégré décrite par Ebrahim & Rangan.

Les investisseurs à impact occupent ainsi une position intermédiaire : trop proches des entreprises individuelles pour avoir une vue systémique sur les chaînes causales, mais contraints de les comparer entre elles pour éclairer leurs décisions d'investissement.

2.1.2 Vers des indicateurs spécifiques aux activités des entreprises

Pour résoudre le problème de la comparabilité des indicateurs d'impact social, plusieurs auteurs ont proposé de converger vers un socle commun d'indicateurs standardisés (Ferreiro & Kölbel, 2025 ; Crucke & Decramer, 2016). Dès 2014, le Comité Français sur l'investissement à impact social appelait à un cadre de mesure commun. La standardisation présenterait un avantage supplémentaire : limiter le biais de sélection qui conduit à ne retenir que les impacts les plus favorables (Arjaliès et al., 2018 ; Kaufmann et al., 2025).

L'analyse d'Ebrahim & Rangan (2014) soutient cependant que les entreprises devraient se concentrer sur la mesure de ce qu'elles font directement, en lien avec leur mission opérationnelle, plutôt que de chercher à s'attribuer des résultats issus de chaînes causales plus larges.

Observations de terrain

Si la littérature académique posait cette logique principalement au niveau des entreprises, nos observations montrent qu'elle s'applique également au niveau des investisseurs, qui sélectionnent et co-construisent souvent avec les entreprises des indicateurs ancrés dans leurs activités directes, plutôt que dans des résultats finaux qui leur échappent partiellement.

Prenons l'exemple d'une plateforme d'orientation professionnelle en ligne, qui aide des apprenants à clarifier leur projet et à identifier des voies possibles. Le retour à l'emploi peut ensuite dépendre d'une formation complémentaire, d'un accompagnement par

d'autres structures, de la conjoncture du marché local — autant de facteurs qui échappent à la plateforme. Son intervention constitue une étape intermédiaire vers l'emploi mais ne peut seule en garantir l'aboutissement. Un indicateur usuel tel que le retour à l'emploi apparaît donc peu adapté pour évaluer l'impact de cette entreprise : à ce sujet, un expert de l'impact a souligné « la difficulté à démontrer l'impact et la causalité, dans la mesure où l'intervention ne constitue qu'une étape parmi d'autres sur le chemin vers l'emploi ». Des experts consultatifs ont donc recommandé de mesurer plutôt l'amélioration de la confiance en soi des apprenants face à leur parcours professionnel, niveau auquel l'entreprise peut revendiquer une contribution davantage isolable.

Cette logique est renforcée par la littérature sur la mesure d'impact social : face à la diversité des entreprises à impact — secteurs hétérogènes, géographies différentes, intentionnalités variées (Höchstädter & Scheck, 2015) — la mesure doit être adaptée à la taille, au secteur et au degré d'innovation sociale de chaque organisation (Mulgan, 2010 ; Nicholls, 2009), et permettre un pilotage effectif articulé avec les activités concrètes des entreprises (Leclerc & Sibieude, 2025).

Cette logique a cependant un coût direct sur la comparabilité des mesures.

2.1.3 Deux freins en cascade : l'absence d'objectifs comparables et la difficulté de hiérarchisation

Le recours à des indicateurs spécifiques produit deux freins à la comparabilité et à la consolidation d'impacts sociaux. Tout d'abord, l'absence d'indicateur standardisé rend difficile la définition d'un niveau de référence établi selon un principe commun. Kroeger & Weber (2014) proposent par exemple d'utiliser, comme seuil de référence, la moyenne de satisfaction déjà atteinte par les autres individus dans une région donnée pour un domaine de vie. Le seuil reste ainsi local, mais il repose sur un principe comparatif : ce que d'autres ont atteint. Cette approche a l'intérêt de traduire indirectement la difficulté du contexte — dans une région où le niveau de satisfaction est structurellement faible, une amélioration modeste y représente une progression relativement importante. Les méthodologies de salaire ou de revenu décent reposent sur un principe différent : le seuil y est également défini localement, mais de façon normative, à partir des ressources jugées nécessaires pour vivre décemment, indépendamment du niveau atteint par les autres. Cette solution s'effondre dès lors que les indicateurs deviennent spécifiques : il n'existe alors ni référentiel normatif établi, ni niveau comparatif disponible, permettant de définir quel serait, par exemple, un niveau idéal de confiance en soi vis-à-vis de son orientation professionnelle.

Observations de terrain

En pratique, les investisseurs définissent parfois des cibles d'impact à l'échelle de chaque entreprise, ce qui permet de mesurer un taux d'atteinte des cibles agrégé. Mais ces cibles sont souvent fixées au cas par cas, notamment à partir des résultats historiques et

d'hypothèses permettant de les projeter dans le futur. Or ces hypothèses sont elles-mêmes plus ou moins conservatrices ou ambitieuses, sans qu'un principe commun extérieur ne vienne les encadrer — ce qui aboutit à des exigences variables d'une entreprise à l'autre.

Comparer un 80 % d'atteinte fondé sur un objectif facile à un 60 % fondé sur un objectif ambitieux ne dit rien sur la valeur sociale relative produite.

Deuxièmement, même si deux entreprises atteignent de bons niveaux sur leurs indicateurs respectifs, il est difficile de hiérarchiser leurs impacts sans un référentiel de valeur partagé. Là où Kroeger & Weber travaillaient avec des domaines de satisfaction génériques dont l'importance relative est supposée comparable, on se retrouve ici avec des indicateurs très hétérogènes dont l'importance respective n'est pas établie. Un niveau satisfaisant d'amélioration de la santé vaut-il un niveau satisfaisant d'amélioration du niveau en langue vivante ? La question n'a pas de réponse technique, elle est normative.

Cette difficulté a une autre face : la comparabilité des indicateurs et la comparabilité des bénéficiaires sont liées. Si deux besoins sont comparables en nature, un faible niveau de satisfaction de ces deux besoins indique une vulnérabilité similaire des personnes ciblées, et la comparaison entre entreprises a du sens. Mais si les besoins sont fondamentalement différents, les personnes dont ces besoins restent insatisfaits ne sont pas dans la même situation : une personne en situation de sous-nutrition et une personne sans accès à des cours de langue vivante ne font pas face au même degré de vulnérabilité. Faute de hiérarchisation des besoins adressés, la comparaison des bénéficiaires ciblés n'a pas de fondement.

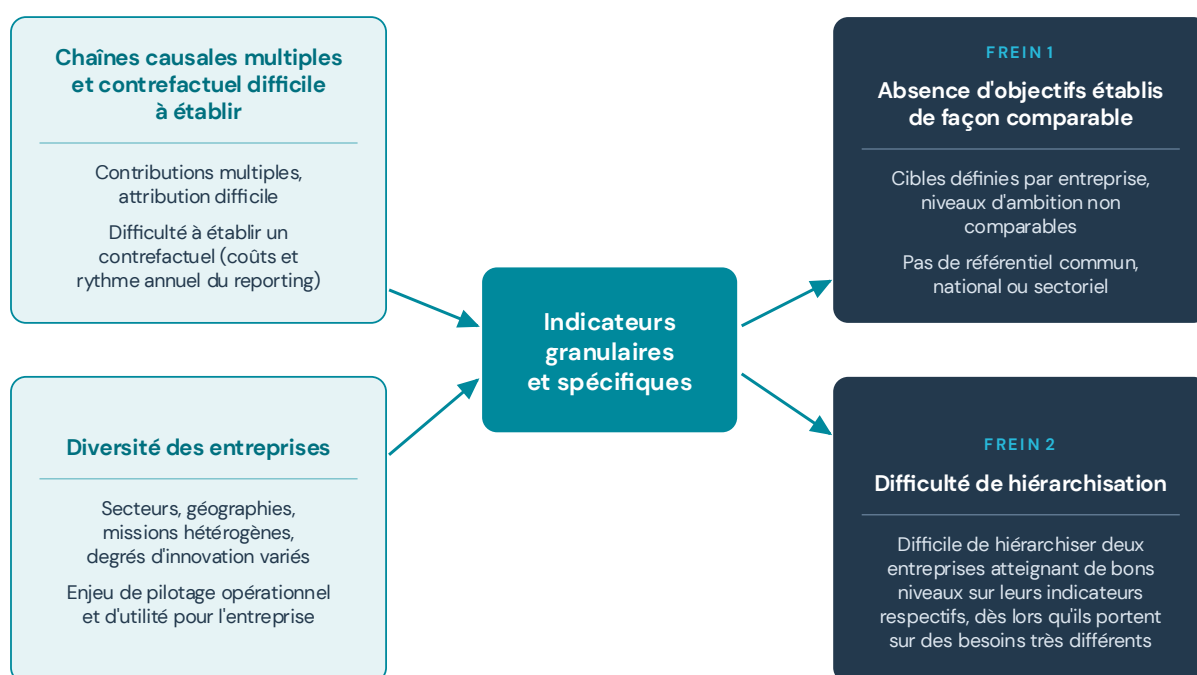


Figure 2. Freins 1 et 2 : L'absence d'objectifs comparables et la difficulté de hiérarchisation des impacts

2.2 L'hétérogénéité des données

La littérature insiste sur la nécessité de disposer de mesures fiables, pertinentes et comparables (Arjaliès et al., 2020) — une exigence rarement satisfaite en pratique. Les approches disponibles sont multiples, qualitatives ou quantitatives, expérimentales ou participatives (Leclerc & Sibieude, 2025), et l'utilisation de méthodologies différentes pour mesurer un même objet peut donner lieu à des résultats différents.

Observations de terrain

Un même objet d'étude — le lien social, par exemple — peut être mesuré par des interactions observées, par des échelles subjectives administrées lors de questionnaires, ou par des indicateurs comportementaux indirects comme la fréquence de participation à des activités collectives. Ces méthodes ne mesurent pas exactement la même chose, et leurs résultats ne sont pas directement comparables.

« On peut évaluer l'évolution d'un indicateur comme la confiance en soi de façon qualitative, en demandant à la personne comment elle perçoit qu'elle a évolué. Certains nous disent que ce n'est pas valable car c'est subjectif. Mais ce qui est important, c'est justement cette perception. Par exemple, quand on mesure l'isolement d'une personne (objectif), ce n'est pas la même chose que le sentiment de solitude (perception). » —
Cabinet de conseil, Expert de l'impact

La fiabilité des données varie également selon l'ampleur des études : certaines se limitent à une enquête de quelques minutes, d'autres mobilisent plusieurs années d'observation (Becker, 2001). En outre, les données mobilisées pour mesurer l'impact social sont rarement homogènes. Dans certains cas, les investisseurs sont confrontés à un manque de données (Arjaliès et al., 2018). Les données primaires collectées auprès des bénéficiaires coexistent fréquemment avec des données secondaires extrapolées d'études existantes, menées dans d'autres contextes, sur d'autres populations. Les données secondaires reposent cependant sur des hypothèses de transférabilité rarement explicitées. Comme le soulignent Pawson & Tilley (1997) avec le concept d'évaluation réaliste, les effets d'une activité dépendent du contexte dans lequel elle s'inscrit et des mécanismes qu'elle y active. L'impact n'est pas automatiquement transférable d'un terrain à un autre. Ainsi, lorsque données primaires et estimations secondaires sont présentées côte à côte sans distinction, deux entreprises peuvent afficher des résultats similaires sur des fondations très inégales, questionnant ainsi la comparabilité.

Sur l'impact social, on n'explicite pas assez ce qui relève de la mesure d'impact et ce qui est de l'estimation. On met ce qu'on suit côte à côte avec l'estimation. —
Asset manager, Expert de l'impact

”

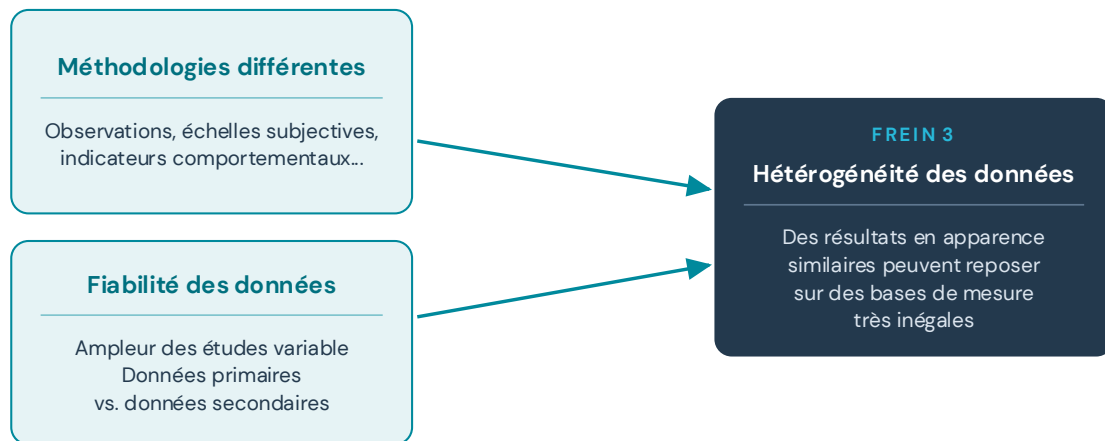


Figure 3. Frein 3 : L'hétérogénéité des données.

2.3 Un enjeu supplémentaire si l'on souhaite agréger : la question de la fongibilité des impacts sociaux

La comparaison de deux impacts distincts se complique encore lorsque l'on cherche à les agréger en un chiffre unique représentant l'impact global d'une entreprise ou d'un portefeuille. Pour certains impacts environnementaux, la fongibilité est souvent admise : une tonne de CO₂ qui serait séquestrée ou évitée quelque part peut se mettre au regard d'une tonne de CO₂ émise ailleurs. Sur le plan social, cela soulève de nouvelles questions.

Premièrement, entre impacts sociaux positifs et négatifs : Mattingly & Berman (2006) montrent que ces deux types d'impacts sont empiriquement et conceptuellement distincts. Les premiers traduisent des initiatives volontaires et proactives (ex. formation des employés, prévention de la pollution), tandis que les seconds reflètent des comportements problématiques (ex. conflits sociaux, procès, accidents environnementaux). Une entreprise peut donc cumuler des comportements positifs et négatifs sans que les uns ne compensent ou effacent les autres. Par conséquent, agréger ces dimensions masquerait des facteurs qui ne se compensent pas selon les auteurs.

Observations de terrain

Lors d'un groupe de travail réunissant des praticiens de l'investissement à impact autour des approches de monétisation des impacts sociaux, un intervenant a présenté des méthodologies permettant d'exprimer les impacts en valeur monétaire : d'un côté, une valeur euro de remédiation, répondant à la question « combien cela coûterait-il de réparer le préjudice causé aux parties lésées ? » ; de l'autre, une valeur euro de bien-être, répondant à la question « pour obtenir la même amélioration du bien-être ressenti, quel revenu supplémentaire aurait-il fallu leur donner ? ».

En conclusion de sa présentation, l'intervenant a posé la question de savoir si ces deux types de valeurs pouvaient être comparées et se compenser mutuellement. Sa réponse était négative : « They are in euros but different euros » — bien que libellées dans la même unité, ces valeurs recouvrent des réalités fondamentalement différentes, l'une étant négative (remédiation à un préjudice) et l'autre positive (amélioration du bien-être), affectant des groupes de personnes distincts selon lui. Elles ne peuvent donc pas être agrégées en un bénéfice net sans masquer ce qui les distingue.

Cette position a suscité des échanges dans la salle, certains praticiens soulignant que l'agrégation était précisément l'objectif de la monétisation des impacts.

Dans un tout autre contexte, un expert de l'impact a également exprimé des questionnements au sujet de la compensation entre des impacts sociaux :

« Pour une entreprise qui fait de la réparation d'électroménager à domicile, il s'agit de vérifier que les gens qui se déplacent n'auront pas un impact négatif supérieur à l'impact positif de la réparation. On regarde alors le ratio de tCO2 par rapport à une nouvelle machine. Sur des sujets sociaux ça marche moins : sauver des vies pour compenser d'autres morts ça ne marche pas. » — Asset manager, Expert de l'impact

La question se pose également entre deux impacts positifs. Admettre que l'amélioration du bien-être physique d'une personne âgée et l'amélioration de son lien social sont additionnables, c'est supposer qu'un gain sur l'un peut compenser un manque sur l'autre. De même, agréger dans un score unique l'impact d'une entreprise sur la santé des femmes et celui d'une autre sur l'insertion professionnelle de jeunes en décrochage suppose d'établir un lien de compensation entre deux réalités fondamentalement différentes.

Cette indétermination ne concerne d'ailleurs pas seulement des impacts de nature différente ou de signe opposé, elle est présente jusque dans l'addition d'occurrences d'un même indicateur. Additionner le nombre de personnes formées, ou de bénéficiaires accompagnés, revient à traiter chaque personne comme une unité interchangeable, équivalente à toutes les autres. Or, même à profils comparables et dans un même contexte, chaque situation individuelle recouvre une réalité singulière que le total efface.

Cette question de la fongibilité n'est pas neutre et technique. Elle repose sur des débats normatifs non résolus, qui font obstacle à une agrégation consensuelle des impacts sociaux.

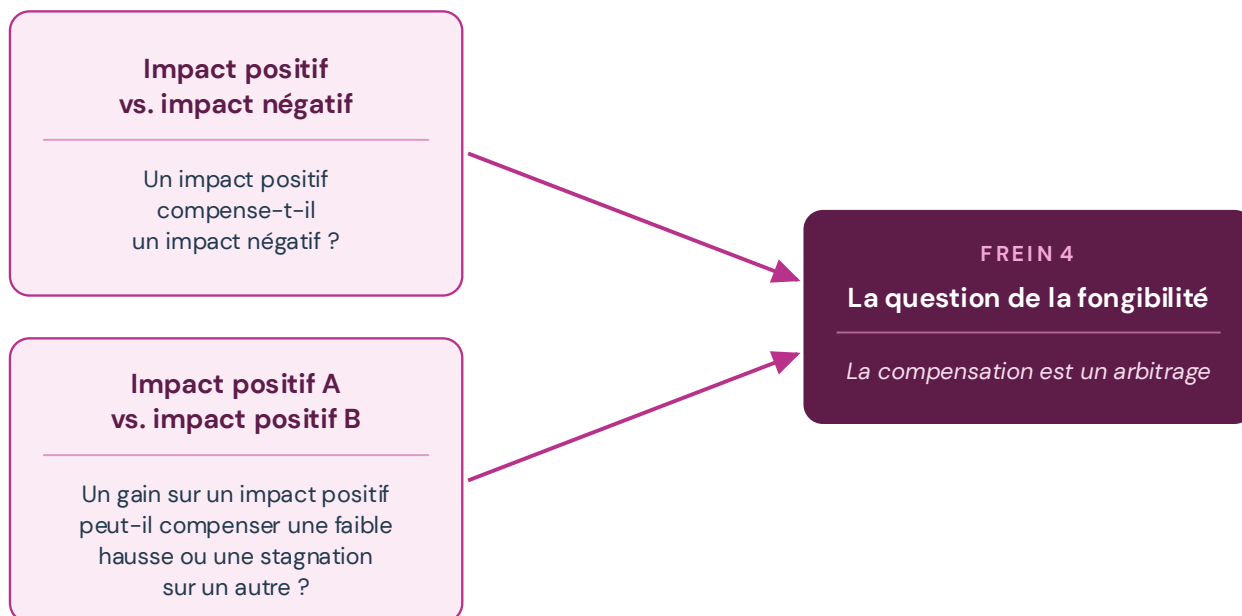


Figure 4. Frein 4 : La question de la fongibilité des impacts sociaux.

Le schéma suivant synthétise l'articulation entre ces quatre freins :

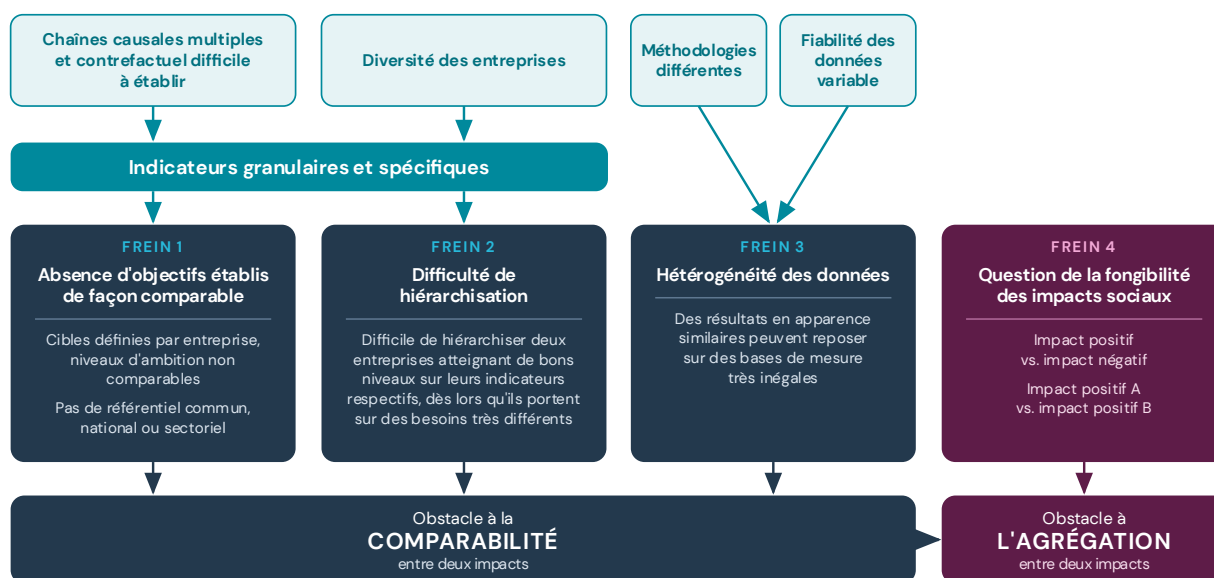


Figure 5. Synthèse : Les freins à la comparaison et à l'agrégation de deux impacts sociaux

3. Comparer deux entreprises : un obstacle supplémentaire

Les difficultés exposées au niveau de la comparaison de deux impacts distincts s'accumulent lorsque l'on cherche à comparer deux entreprises entre elles. S'y ajoute un obstacle spécifique à ce niveau d'analyse : l'hétérogénéité des périmètres de mesure.

3.1 Un obstacle supplémentaire : l'hétérogénéité des périmètres de mesure

Les évaluations d'impact sont exposées à une complexité computationnelle parce que de nombreux éléments peuvent être pris en compte (Stievenart & Pache, 2014). Une démarche de mesure d'impact implique de prioriser des informations par rapport à d'autres et les résultats sont ainsi nécessairement incomplets. Cela induit donc inévitablement un biais.

Observations de terrain

Une entreprise opérant dans le secteur du soin à la personne impacte de nombreuses parties prenantes : les bénéficiaires directs bien sûr ; mais aussi leurs proches et aidants, dont la charge peut être allégée par l'intervention ; les professionnels de santé partenaires, que l'entreprise accompagne ou complète dans leurs activités ; les communautés locales, si l'entreprise a un ancrage territorial ; ou encore les pouvoirs publics concernés par les coûts évités que l'intervention peut générer.

Les exercices de mesure étant intensifs en ressources et en temps, il n'est pas possible de couvrir l'ensemble des parties prenantes dans le cadre d'une telle démarche. Un périmètre est alors priorisé. Celui-ci peut être défini en fonction de plusieurs facteurs variables : l'intentionnalité de l'entreprise (quels effets cherche-t-elle délibérément à générer, au cœur de sa stratégie ?), la matérialité perçue des enjeux (quelles parties prenantes sont les plus affectées ?) et les choix de communication. Il peut également évoluer dans le temps, s'étendant au fur et à mesure que l'entreprise gagne en maturité sur la mesure.

« Nous on a choisi de cibler les bénéficiaires. Il y a plein de parties prenantes, mais on a fait le choix dès le début de collecter les impacts pour cette partie prenante parce qu'on voulait vraiment arriver à comprendre et à valoriser la valeur ajoutée de nos contenus, formats, intervenants, pédagogie mise en œuvre, etc. » — Expert de l'impact, Entreprise

« On n'a pas lancé de mesure sur les responsables d'animation parce qu'on ne peut pas tout faire à la fois, mais évidemment nos services leur offrent un portefeuille d'activités plus important. » — Expert de l'impact, Entreprise

Ce frein ne tient d'ailleurs pas uniquement à l'hétérogénéité des indicateurs, il se pose également entre indicateurs homogènes. Deux entreprises qui rapportent un même indicateur — par exemple le nombre d'emplois soutenus — ne sont pas directement comparables si l'une ne

comptabilise que ses salariés directs quand l'autre y inclut les emplois indirects de sa chaîne de valeur.

Ce problème est analogue, dans le domaine climatique, à la distinction entre les scopes 1, 2 et 3 des émissions de gaz à effet de serre. Une entreprise qui ne mesure que ses émissions directes n'est pas directement comparable à une entreprise qui inclut ses émissions indirectes liées à sa chaîne d'approvisionnement ou l'utilisation en aval des produits qu'elle vend : le périmètre choisi change fondamentalement ce qu'on mesure.

Sans un cadrage commun du périmètre de mesure, les écarts observés entre entreprises peuvent ainsi refléter des différences réelles d'impact, ou simplement des choix différents de périmètre.

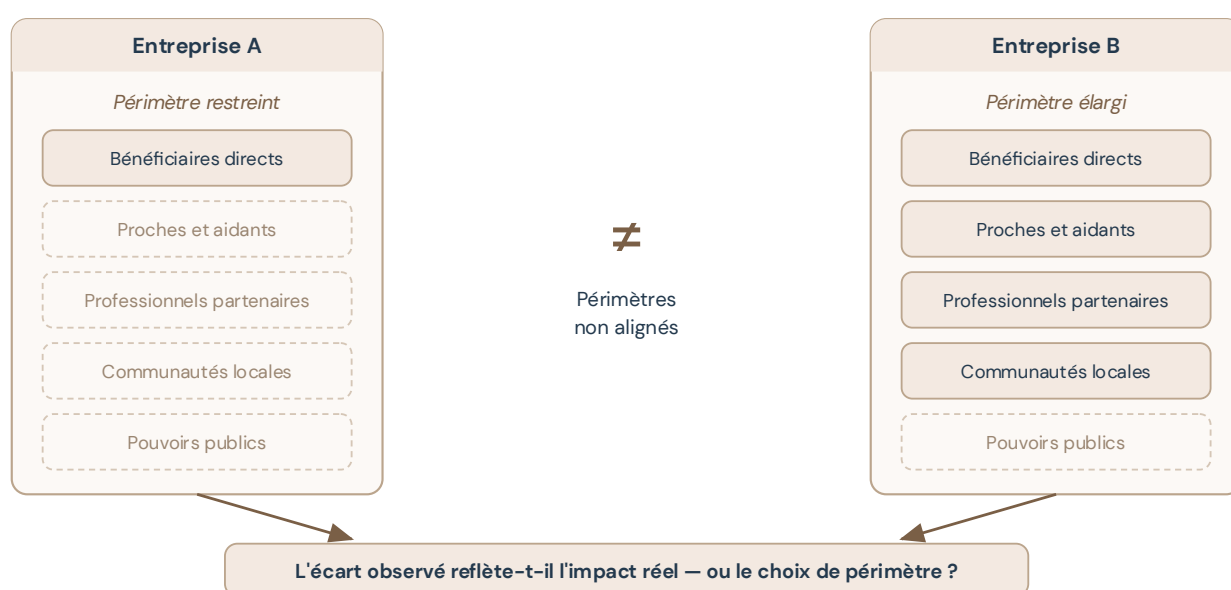


Figure 6. Frein 5 : L'hétérogénéité des périmètres de mesure

3.2 La comparaison inter-entreprises : une accumulation de tous les obstacles

La comparaison inter-entreprises cumule l'ensemble des obstacles identifiés au niveau des impacts individuels — indicateurs granulaires et spécifiques aux entreprises qui ne permettent pas d'accéder à des seuils communs et qui complexifient la hiérarchisation ; hétérogénéité des méthodes et des données — et y ajoute celui du périmètre. Même deux entreprises opérant dans le même secteur avec des intentionnalités d'impact proches peuvent difficilement être comparées directement si leurs indicateurs sont différents, leurs méthodes de collecte divergentes et leurs périmètres non délimités selon les mêmes principes.

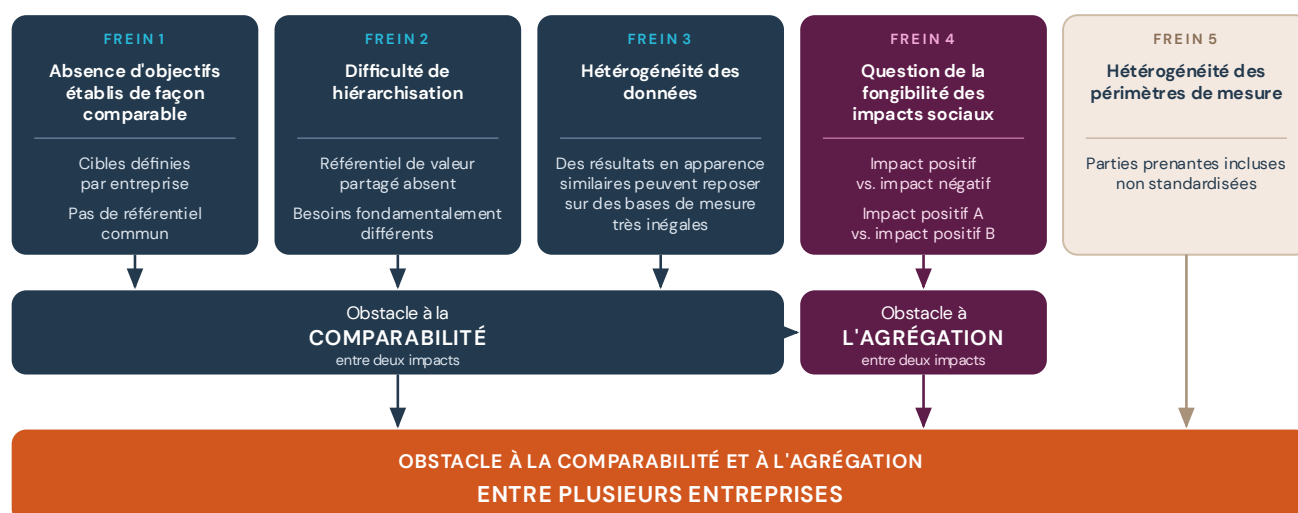


Figure 7. Synthèse : Les freins à la comparaison et à l'agrégation des impacts sociaux entre plusieurs entreprises

Pour l'investisseur à impact, qui doit pourtant prendre des décisions d'allocation entre ces entreprises, cette situation est exigeante. Elle appelle des approches qui cherchent à construire des référentiels permettant de situer les impacts les uns par rapport aux autres, afin d'éviter de perdre en granularité par le biais de la standardisation.

4. Conclusion : pistes pour une mesure d'impact social comparable

Ce white paper a exploré les freins structurels qui rendent la comparaison et l'agrégation des indicateurs hétérogènes difficiles, à travers le cas de l'impact social. La tension entre spécificité des indicateurs et besoin de consolidation est au cœur du problème. Les pratiques de consolidation s'appuient aujourd'hui principalement sur des indicateurs de volume — nombre de bénéficiaires, emplois créés — transversaux à des activités très variées. Mais ces indicateurs ne rendent pas compte de la qualité de l'impact ni de la vulnérabilité des populations ciblées. L'enjeu est dès lors de lever ces freins, ce qui appelle plusieurs pistes de travail complémentaires. Il convient toutefois de souligner d'emblée que l'ampleur de ces chantiers dépend étroitement de l'échelle à laquelle on souhaite comparer : comparer au sein d'un fonds, entre fonds d'un même gestionnaire, ou entre gestionnaires suppose des degrés de coordination différents.

Hiérarchiser les impacts ciblés et la vulnérabilité des bénéficiaires. Face à des indicateurs fondamentalement hétérogènes, une première voie consiste à construire des dimensions de hiérarchisation des impacts selon des principes communs : à quel point tel impact est-il considéré comme plus fondamental qu'un autre ? Quelle est la vulnérabilité des populations ciblées ? Ces choix impliquent des arbitrages normatifs qui doivent être rendus explicites et assumés.

Certains investisseurs s'engagent déjà dans cette direction, en cherchant à caractériser la vulnérabilité socio-économique de leurs bénéficiaires. Par exemple, le pourcentage de bénéficiaires relevant d'un dispositif d'aide publique constitue un proxy de leur degré de

vulnérabilité — non pas un indicateur de hiérarchisation directe, mais une tentative de qualifier l'intensité du besoin adressé.

Pour la caractérisation on demande depuis combien de temps la personne est en structure, si elle est veuve, divorcée, seule, afin de cibler la situation familiale. On identifie clairement le lieu de déploiement de l'atelier, si on est en zone blanche très peu pourvue d'activités. On est aussi dans cette réflexion d'identifier un peu mieux les difficultés de ces bénéficiaires, dans quelle mesure on est au plus proche et on répond à ce public-là. Pour l'instant on a cette sensibilité et perception dont on se rend compte mais on n'a pas encore vraiment posé les choses en disant « c'est vraiment ça qui va nous permettre de discriminer des actions les unes par rapport aux autres ». — Expert de l'impact, Entreprise

”

Cette piste reste cependant difficile à opérationnaliser. La vulnérabilité est multidimensionnelle : identifier quelle dimension prime sur les autres pour établir une hiérarchie suppose également des choix normatifs. De même, hiérarchiser des impacts entre eux — selon les besoins qu'ils adressent — peut sembler évident dans certains cas, mais peut devenir très délicat (soins d'urgence versus santé mentale, santé versus éducation). Ces questions normatives n'ont pas de réponse technique.

Définir des objectifs et des seuils selon des principes communs. Une deuxième voie consiste à harmoniser la manière dont les objectifs d'impact sont fixés et mesurés. Lorsque la qualité d'un impact est mesurée à partir d'échelles de Likert¹ administrées aux bénéficiaires, ces échelles sont par définition normalisées et bornées — elles offrent ainsi une base de comparaison délimitée, indépendamment de l'indicateur mesuré. Cette approche constitue un exemple de voie prometteuse pour construire des indicateurs d'impact comparables entre entreprises, sans nécessiter de standardisation des indicateurs eux-mêmes.

Harmoniser les méthodologies de collecte. Une convergence progressive des pratiques — sur les protocoles d'enquête, les échelles utilisées, la distinction entre données primaires et secondaires — permettrait de renforcer la comparabilité des mesures. Ce chantier suppose un travail collectif dont l'ampleur varie selon l'échelle visée : au sein d'un fonds, il implique un alignement entre l'investisseur et ses entreprises ; entre fonds ou entre gestionnaires, il nécessite une coordination de place beaucoup plus large.

Standardiser ou expliciter les périmètres. Définir de façon cohérente et justifier quelles parties prenantes sont incluses dans le périmètre de mesure — et dans quelle mesure ces périmètres peuvent être considérés comme comparables entre eux — permet de s'assurer que les comparaisons reposent sur une base homogène, et non sur des choix de cadrage non explicités.

Enfin, il faut reconnaître les limites de l'exercice. Ce white paper est avant tout un diagnostic : il explique pourquoi la comparaison et l'agrégation des impacts hétérogènes sont difficiles, et

¹ Echelle de mesure psychométrique composée de niveaux de réponse ordonnés (par exemple de « pas du tout d'accord » à « tout à fait d'accord »), utilisée pour quantifier le degré d'accord, de satisfaction ou de perception d'un répondant face à une affirmation donnée.

pourquoi ces freins ne se résolvent pas aisément — l'impact social en constitue un cas extrême, qui cumule l'ensemble de ces difficultés et un degré d'hétérogénéité que l'on retrouve dans une moindre mesure sur d'autres thématiques. Les pistes esquissées ci-dessus atténuent ces freins sans les faire disparaître : une agrégation pleinement rigoureuse des impacts sociaux n'existe sans doute pas. Pour autant, le sujet ne s'arrête pas à ce constat. En pratique, il revient à chaque investisseur de définir une approche pragmatique — nécessairement imparfaite — qui lui permette malgré tout de rendre compte de ses impacts, d'orienter ses décisions et de piloter ses portefeuilles. La véritable question n'est dès lors pas tant de savoir si ces approches sont parfaites, mais de déterminer quelles imperfections et quels biais sont acceptables au regard de l'exercice mené et de l'usage qui en sera fait. La mesure d'impact social implique toujours des choix — sur le périmètre ; sur les indicateurs et leur potentielle agrégation, les pondérations appliquées ; sur les méthodes de collecte — qui influencent les résultats obtenus. Rendre ces choix explicites, et distinguer clairement ce qui relève de la mesure directe et ce qui relève de l'estimation, est une condition de la crédibilité et de l'utilité des mesures d'impact social.

Ces questions s'inscrivent dans un programme de recherche plus large mené au sein du laboratoire i3-CRG, École polytechnique, CNRS, Institut polytechnique de Paris, Palaiseau.

Références

- Arjaliès, D.-L., Bouchet, V., Crifo, P., & Mottis, N. (2020). La mesure d'impact et l'investissement Socialement Responsable (ISR) : Un tour d'horizon. Dans E. Tchotourian, L. Bres, & L. Geelhand de Merxem (Dir.), *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*. 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*.
https://www.economie.gouv.fr/files/files/PDF/RapportSIFce_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*. <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* (1re éd.). 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, E., 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*. ESSEC Business School, Avise, Mouvement Impact France. https://www.avise.org/sites/default/files/atoms/files/20211203/1223361021_petit_precis_de_evaluation_de_limpact_social_2_web_1.pdf
- Stievenart, E., & Pache, A.-C. (2014). Evaluer l'impact social d'une entreprise sociale : Points de repère. *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/>