

Comment estimer le rendement de l'investissement social ?

Arthur Heim



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

Les documents de travail de France Stratégie présentent les travaux de recherche réalisés par ses experts, seuls ou en collaboration avec des experts extérieurs. L'objet de leur diffusion est de susciter le débat et d'appeler commentaires et critiques. Les documents de cette série sont publiés sous la responsabilité éditoriale du commissaire général. Les opinions et recommandations qui y figurent engagent leurs auteurs et n'ont pas vocation à refléter la position du Gouvernement.

Table des matières

Introduction	3
1. Les évaluations de l'investissement social	5
1.1. Une dominante macroéconomique dans les évaluations existantes.....	5
1.1.1. Des classements de dépenses.....	5
1.1.2. Les liens entre dépenses sociales et performances macroéconomiques....	7
1.1.3. Les limites des approches macroscopiques et internationales.....	9
1.2. Quelles preuves pour l'investissement social ?	11
1.3. Évaluation d'impact et investissement social.....	14
1.3.1. Les difficultés de l'évaluation d'impact.....	14
1.3.2. Quelques évaluations d'impact de politiques d'investissement social	18
2. Des effets de l'investissement social à son rendement	30
2.1. Les approches hypothétiques	31
2.2. L'apport de l'évaluation socioéconomique appliquée aux politiques sociales	32
2.3. Des analyses coûts-efficacité aux analyses coûts-bénéfices.....	34
2.3.1. Définir le ratio coût/efficacité.....	34
2.3.2. De l'efficacité au bénéfice : l'inflation des hypothèses	38
2.3.3. Quelques analyses coûts-efficacité et coûts-bénéfices associées à des évaluations	41
2.3.4. Gérer l'incertitude des impacts et des valeurs par réplication	44
2.4. Retours sociaux sur investissement ou analyses coûts-bénéfices ?.....	45
3. Quelles évaluations pour l'investissement social en France ?	47
3.1. Pertinence et cohérence des politiques sociales au regard de l'investissement social.....	47
3.2. Le recours aux modèles théoriques, calibrations et simulations	49
3.3. Recenser les évaluations d'impact existantes	50
3.4. Les données inexploitées.....	52
3.5. Utiliser l'expérimentation	54
3.6. Identifier le rendement d'investissements sociaux.....	55
Conclusion	56
Bibliographie	58

Résumé

L'investissement social apparaît à la fin des années 1990 comme un modèle d'action sociale renouvelé pour faire face à l'évolution de sociétés fragilisées par l'émergence de nouveaux risques sociaux. Il consiste à investir dans le capital humain des individus de façon à augmenter et à maintenir leurs compétences et leur capacité à supporter les chocs. La notion s'appuie en outre largement sur la promesse que la dépense sociale d'aujourd'hui rapportera à un moment donné. Sa crédibilité repose donc sur l'existence de travaux permettant d'identifier ce rendement.

Ce document de travail analyse les recherches existantes sur les impacts de l'investissement social. Il montre que l'approche consistant à lier les dépenses d'investissement social aux performances macroéconomiques des États n'est pas en mesure d'identifier de façon crédible l'effet causal de l'investissement social et *a fortiori* son rendement.

Ce document propose en outre d'adopter une vision plus locale et de séparer en deux étapes l'identification du rendement de l'investissement social. La première consiste à recourir à des évaluations d'impact en adoptant des méthodes qui permettent d'interpréter les résultats comme un lien de cause à effet. Une littérature scientifique abondante évalue déjà l'effet de politiques assimilables à de l'investissement social : la collecte de ces informations est un prérequis pour engager de nouvelles réformes d'investissement social. Nous proposons d'adopter une démarche de collecte systématique et d'analyses statistiques, en produisant davantage de méta-analyses.

La deuxième étape consiste à intégrer les résultats des évaluations d'impact dans des modèles de calculs socioéconomiques, afin de comparer les bénéfices des investissements ainsi identifiés aux coûts réels et sociaux de leur mise en œuvre. Le succès de cette étape suppose toutefois de surmonter deux obstacles importants :

- améliorer les données disponibles sur le coût des investissements sociaux mais aussi trouver la valeur monétaire des effets induits par l'investissement social ;
- prendre en compte la grande incertitude des estimations d'impact et de coût.

Mots clés : investissement social, rendement, évaluation, impact, politiques publiques

Comment estimer le rendement de l'investissement social ?

Arthur Heim¹

Introduction

Depuis la fin des années 1990, l'investissement social apparaît comme un modèle d'action sociale renouvelé (Midgley, 1999). Centré sur l'accompagnement des individus tout au long de leur parcours de vie, il se veut une réponse à l'évolution des risques sociaux fondée sur la prévention.

L'investissement social (IS par la suite) consiste à « investir dans les personnes en adoptant des mesures pour renforcer leurs compétences et leurs capacités, leur autonomie et leur permettre de participer pleinement au monde du travail et à la société » (Commission européenne, 2015). Il s'inscrit donc dans une logique d'action publique focalisée sur la dotation des individus en capacités de production, d'adaptation et de résilience face aux risques sociaux. L'IS doit viser, par ailleurs, et en particulier, les populations les plus défavorisées, joignant ainsi les théories de la justice sociale de Rawls et des capacités de Sen (Delors & Dollé, 2009). On comprend alors que l'investissement social peut se voir attribuer d'autres finalités, notamment celles de concilier des objectifs économiques avec un objectif social (maintien de la cohésion sociale par la lutte contre les inégalités) et démocratique (permettre l'exercice d'une citoyenneté active).

Parce qu'il apparaît à la fois comme un nouveau paradigme économique et politique, un ensemble de « politiques sociales productives » (Hemerijck A., 2012) et une stratégie d'action publique, le champ de l'investissement social n'est pas bien défini. Une partie importante de la littérature s'est attachée à établir les contours de la notion dans ces trois dimensions en distinguant des dépenses actives et passives, des risques anciens et nouveaux, etc. (Morel, Palier & Palme, 2013). Des domaines consensuels d'action émergent autour des politiques de la petite enfance, d'éducation, des politiques d'insertion et d'emploi, de conciliation vie familiale/professionnelle, de la formation tout au long de la vie. Au contraire, d'autres sont souvent exclues du champ : la santé et la retraite notamment.

¹ Département Société et Politiques sociales, France Stratégie, et Direction des Statistiques, des Études et de la Recherche, Caisse nationale des allocations familiales.

De plus, dans un contexte où les ressources publiques sont rares et déjà largement engagées, les marges de manœuvres pour des réformes des politiques sociales sont minces. Cette contrainte pousse souvent les chercheurs et décideurs à produire des analyses à budget constant ou réduisant les dépenses, instrumentalisant parfois les évaluations des politiques publiques. Pour Jean Pisani-Ferry, « l'évaluation ne doit pas être perçue uniquement comme un instrument de la réduction de la dépense » car c'est prendre le risque d'exclure des stratégies prometteuses d'action publique mais qui peuvent nécessiter un surcoût à court terme.

C'est là que la notion d'investissement social prend tout son sens et se distingue d'une analyse plus spécifique d'évaluation de politiques publiques ou d'économie publique.

Toutefois, cette approche est controversée sur les effets qu'elle peut produire et la fiabilité du retour sur investissement. En effet, ses résultats peuvent être très diffus, n'apparaître qu'à long terme et être difficilement identifiables, notamment en raison de l'articulation de différentes politiques sociales (Palier, 2014). De plus, ce qui apparaît dans l'évaluation de l'investissement social comme des preuves empiriques varie de « résultats évocateurs ou anecdotiques à des expérimentations aléatoires contrôlées, focalisées sur peu ou beaucoup de pays ou individus, utilisant des méthodes et philosophies variées pour faire apparaître des tendances, des écarts ou approcher la causalité ». En conséquence, les chercheurs et les décideurs politiques se fondent sur des informations de qualité très différente ce qui nuit à la crédibilité et à la pertinence de la notion.

Bonoli (2012) rappelle que la notion d'investissement social est largement basée sur la promesse que l'argent dépensé aujourd'hui rapportera à un moment dans le futur. La crédibilité de cette démarche dépend de l'existence de travaux empiriques capables de démontrer ces retours.

C'est pourquoi l'investissement social doit se doter d'un champ évaluatif performant afin de faire émerger des éléments probants et d'asseoir son utilité, parfois instrumentalisée pour légitimer les politiques d'activation et promouvoir le rendement des programmes intensifs d'éducation en bas âge. Pour autant, l'investissement social n'est-il que la somme de dispositifs évalués et probants ?

Ce document de travail propose des perspectives d'évaluation pour l'investissement social en débutant par une discussion sur les travaux existants, distinguant les approches globales des évaluations plus fines. Il mise sur une analyse en deux temps : une évaluation de l'effet de l'investissement social – qui peut être expérimentale, quasi expérimentale, ou tiré de synthèses de littérature, mais pouvant *in fine* être interprété comme un lien de cause à effet – puis, une évaluation du rendement – compris comme des calculs socioéconomiques rapportant les effets précédemment évalués aux coûts de l'intervention.

1. Les évaluations de l'investissement social

La démonstration de la rentabilité de l'investissement social mobilise des analyses aux méthodologies extrêmement variées. Les travaux précurseurs ont principalement concerné les évaluations de l'efficacité des interventions précoces auprès des enfants défavorisés en mobilisant aussi bien des travaux fondés sur des expérimentations contrôlées et des analyses économétriques sophistiquées (les travaux d'Heckman principalement) et des analyses plus politiques comme celles du sociologue Esping-Andersen. Au-delà de cette approche, l'évaluation de l'IS a largement consisté à analyser et comparer les dépenses sociales des pays dans une perspective macroéconomique en mettant en relation ces comparaisons avec des résultats agrégés en termes de chômage, de croissance, de taux d'emploi, de taux d'emploi à forte valeur ajoutée, etc.

1.1. Une dominante macroéconomique dans les évaluations existantes

1.1.1. Des classements de dépenses

Les travaux sur l'investissement social ont principalement cherché à identifier les ensembles de politiques sociales qui peuvent être qualifiés d'investissements sociaux, et ensuite à expliquer les performances économiques mais aussi sociétales d'un État en fonction de l'effort d'investissement. Ces travaux ne se revendiquent généralement pas comme des évaluations mais cherchent plutôt à supporter la thèse selon laquelle les pays développés ont déjà amorcé la transition vers différentes formes d'État d'investissement social (Palier, 2005 ; Cantillon & Lancker, 2011). Différents modèles peuvent ainsi être identifiés et comparés tant sur l'effort d'investissement social que sur les performances macroéconomiques. L'IS offre alors un cadre d'analyse pour caractériser les dépenses sociales d'investissement et les autres.

De Deken (2016) considère deux méthodes pour mesurer l'investissement social dans une perspective internationale : la première cherche à mesurer l'impact de la protection sociale en élaborant des indicateurs d'efficacité, la seconde décrit l'évolution des caractéristiques du système de protection sociale en examinant la dépense sociale, en la ventilant et en décrivant ses variations.

Plusieurs travaux analysent ainsi la composition des dépenses entre les pays ou leur évolution dans le temps et évaluent leur efficacité. Ils proposent des classifications pour séparer les dépenses d'investissement des autres. Nikolai (2012) utilise les données de dépenses sociales collectées par l'OCDE pour séparer les dépenses de compensation des dépenses d'investissement. Vandenbroucke & Vleminckx (2011) exploitent les mêmes données mais fondent leur classification sur la notion de dépenses anciennes ou nouvelles. Cette notion est reprise par Vaalavuo (2013) dans une analyse des effets redistributifs des dépenses. De Deken (2013) élabore une classification plus flexible autour des travaux de Vandenbroucke et Vleminckx et propose des sous-catégories

combinant les notions de compensation *versus* investissement et de risques anciens *versus* nouveaux.

Figure 1 – Classification des dépenses compensatoires et d'investissement social d'après De Deken (2013)

Table 1. Linking social investment to identifiable social policy categories.

Compensatory Policies	Social Investment Policies		
Compensating exit of labour market participants	Maintaining or restoring capacity of labour market participants	Facilitating entrance of new labour market participants	Investing in the capacity of new labour market participants
<ul style="list-style-type: none"> - Old Age Pensions - Survivor Pensions - Incapacity Benefits (LT) - Early Retirement - Unemployment Assistance - Social Assistance 	<ul style="list-style-type: none"> - Unemployment Insurance (ST) - Maternity and Parental leave (ST) - Reintegration of work incapacitated - Active Labour Market Policies 	<ul style="list-style-type: none"> - Care for Elderly and Frail - Child care - Active Labour Market Policies 	<ul style="list-style-type: none"> - Active Labour Market Policies - Child care - Education <ul style="list-style-type: none"> * pre-primary education * primary education * lower secondary education * upper secondary education
	<ul style="list-style-type: none"> - Maternity Leave (LT) - Unemployment Insurance (LT) - Housing benefits - Sickness Pay - Health Care 		

Ces classifications ont l'avantage de fournir un cadre d'analyse simple et explicite et de permettre des comparaisons internationales des dépenses d'investissement social et de leur évolution pour caractériser les différents modèles. Cependant, on peut s'interroger sur le sens et la robustesse de la distinction entre dépenses d'investissement social et les autres formes de dépenses sociales. Comme le note assez justement Nolan (2013), la notion d'investissement social n'est pas privée de sens politique et traduit l'idée que certaines dépenses sociales améliorent les capacités productives des individus et de l'économie dans sa globalité – ce qui n'apparaît pas dans les mouvances libérales ou keynésiennes.

Elle oriente aussi l'attention sur la contrepartie des dépenses sociales en identifiant les produits correspondant à la charge (Damon, 2015).

Cependant, la distinction n'a de sens que si les autres dépenses sociales ne produisent pas (ou très peu) ces effets, ne sont en réalité que des dépenses de consommation, et qu'il est possible de distinguer réellement ces dépenses. En pratique, cette tâche est presque impossible. Par exemple, les prises en charge de longue durée de personnes malades, âgées ou invalides, qui sont généralement considérées comme des dépenses passives ou anciennes, peuvent permettre le retour à l'emploi de personnes qui, en leur absence, auraient été contraintes de s'occuper de celles dans le besoin. L'assurance chômage permet de maintenir le capital humain des individus et laisse le temps à la recherche d'emploi jusqu'à ce qu'un bon « match » soit trouvé. Dit de cette façon, il est presque impossible d'exclure *a priori* des politiques sociales du champ de

l'investissement social car on trouve toujours une façon de dire qu'une politique au moins préserve le capital humain des personnes qui en bénéficient. On se retrouve alors dans une situation où l'investissement social est réduit à un « outil de stratégie politique, voire un élément de rhétorique stratégique » (Damon, 2015).

Ces difficultés de classification des dépenses sont contournées par Hemerijck (2014) qui passe par les fonctions des dépenses pour :

- améliorer le stock de capital humain (fonction stock) ;
- faciliter les transitions sur le marché du travail et tout au long du cycle de vie (fonction de flux) ;
- limiter l'effet des risques sociaux lorsqu'ils se réalisent en produisant des filets sociaux universels de sécurité et des amortisseurs économiques (fonction de *buffer*).

Cette classification paraît plus robuste aux critiques empiriques et normatives qui ont visé celles précédemment décrites ; mais elle introduit des dimensions plus qualitatives et subjectives, qui la rendent moins opérationnelle. Hemerijck & Vydra (2016) proposent de caractériser les investissements sociaux tout au long du cycle de vie en les illustrant par ces fonctions. Par exemple, la fonction stock de l'investissement social en petite enfance est : « Une éducation préscolaire universelle et de qualité favorise le développement cognitif et l'intégration sociale. Elle limite aussi la reproduction générationnelle de la pauvreté ». Celle de flux est « Une éducation préscolaire universelle et de qualité favorise la coopération entre les parents et les enseignants, ce qui permet une continuité de l'apprentissage ». Enfin, celle d'amortisseur est : « Le fait de vivre dans un foyer stable où le risque de pauvreté est faible favorise une alimentation appropriée et le développement émotionnel ». Cette catégorisation bute ainsi sur les mêmes difficultés de classement puisqu'elle implique une ventilation des dépenses par fonctions, relativement subjective et peu discriminante.

Les travaux sur l'investissement social de la Commission européenne renoncent à classer les dépenses sociales, considérant que les systèmes de protection sociale « remplissent trois fonctions : investir dans le domaine social, assurer une protection sociale et stabiliser l'économie » (Commission européenne, 2015). Ces trois fonctions sont remplies par chaque politique à un niveau plus ou moins important. Dès lors, les classifications de dépenses sociales peuvent paraître trompeuses puisque toute dépense agit à la fois comme investissement, protection et stabilisateur et varie suivant les caractéristiques de conception de cette dernière (conditionnalité, durée, etc.), les spécificités du contexte national (complémentarité à d'autres politiques) et les circonstances dans le temps (durée du cycle économique, trajectoire de croissance).

1.1.2. Les liens entre dépenses sociales et performances macroéconomiques

D'autres travaux ont pris le parti de mettre en relation ces classements de dépenses sociales avec les performances macroéconomiques des pays. On trouve différentes

approches allant de statistiques descriptives, de corrélations, à des modélisations économétriques du type régression de panel.

Morgan (2012) interroge ainsi les dépenses d'accueil des enfants en bas âge et distingue les dépenses en nature des transferts monétaires. Nelson & Stephens (2009) analysent le rôle du système éducatif dans la formation du capital humain et expliquent les différences de performances par les caractéristiques politiques des pays². Dans leur chapitre du livre de Morel *et al.* (2012), ils cherchent à mesurer le lien entre plusieurs variables institutionnelles et l'emploi par des modélisations en panel groupé. Ils concluent que la durée et le montant de l'assurance chômage, les arrêts maladie, les dépenses de garde d'enfants, les dépenses d'éducation, les politiques d'activation du marché du travail et le nombre moyen d'années d'étude expliquent significativement les différences de taux d'emploi entre les pays. Plus spécifiquement, ces caractéristiques sont positivement associées à la création de « bons emplois », entendus comme intensifs en connaissance. Ce type d'analyse et de classification est également utilisé dans les travaux de la Commission européenne qui propose différentes décompositions des dépenses sociales et des mesures de leurs effets dans plusieurs dimensions, souvent représentées par des *radar charts* (Bontout, Maestri, & Vaalavuo, 2015). Le rapport présente également plusieurs corrélations pour illustrer l'efficacité des différentes politiques sociales (écart de taux d'emploi entre femmes avec ou sans jeune enfant et taux d'enfants de moins de 2 ans dans des structures d'accueil formelles, part des prestations familiales dans le PIB et différence de taux de pauvreté avant et après transfert, etc.).

Cantillon et Van Mechelen (2013) s'intéressent à la réduction de la pauvreté en Belgique et argumentent que la protection sociale et les mécanismes de prévention existants n'ont pas permis de réduire la proportion de travailleurs pauvres. Leur travail met en évidence *l'effet Matthieu* des politiques d'investissement social en matière de lutte contre la pauvreté. Alors que l'investissement dans la petite enfance est vu dans le paquet investissement social de la Commission européenne comme l'un des principaux leviers de lutte contre les inégalités, Lancker (2013) interroge le recours aux modes de garde dans les 27 pays européens. Il montre alors que les services de garde d'enfants sont moyennement ou faiblement utilisés et que les familles défavorisées y ont beaucoup moins recours que les familles plus favorisées. Il conclut que dans leur grande majorité, l'accès n'est pas suffisant pour produire les effets macroéconomiques attendus sur l'emploi des mères et le développement des enfants et que des efforts d'action publique importants sont nécessaires pour y parvenir.

Ahn & Kim (2015) modélisent – à partir d'un panel de 15 pays de l'OCDE et des données de 1990 à 2007 – la relation entre le PIB par tête et l'emploi à partir de la dépense sociale (et une série d'autres variables explicatives conventionnelles). Leurs résultats indiquent une association positive et significative entre la part de dépense

² Ce travail est principalement fondé sur la thèse de Nelson (2008). Les performances éducatives moyennes sont expliquées par le modèle politique du pays (social-démocrate, conservateur, chrétien-démocrate, etc.).

sociale de services dans la dépense sociale totale avec le PIB par tête et le taux d'emploi, ce qu'ils interprètent comme un signe d'efficacité de l'investissement social. Au contraire, la dépense totale n'affecte pas le PIB par tête et est négativement associée à l'emploi. Vliet & Wang (2015) adoptent une stratégie similaire de régression de panel groupé pour 15 pays européens de 1997 à 2007 afin de voir si l'augmentation ou la stagnation du taux de pauvreté peut s'expliquer par les modifications de dépenses sociales orientées vers l'investissement social. Leurs conclusions indiquent que les effets négatifs des politiques d'investissement social décrites notamment par Cantillon et Van Mechelen (2013) n'est pas généralisable à un groupe plus large de pays européens. Toutefois, en dehors des pays nordiques, les analyses identifient une relation entre les taux de pauvreté et les modifications de la dépense sociale induites par les nouveaux modèles d'Etat-providence. Les différences entre les différentes spécifications font douter de la robustesse du modèle.

1.1.3. Les limites des approches macroscopiques et internationales

L'approche par les dépenses sociales pose trois problèmes importants (Boisson-Cohen & Collombet, 2016).

D'abord, les agrégats de dépenses ne tiennent pas compte des effets de structure et notamment la pyramide des âges et la composition de la main-d'œuvre. Ainsi, le vieillissement de la population conduit mécaniquement à une hausse de la dépense sociale pour les retraites et les dépenses de santé, expliquant l'évolution des « vieilles » dépenses. Leur utilisation requiert ainsi des précautions supplémentaires qui leur font perdre en lisibilité et rajoutent des hypothèses et des transformations.

Ensuite, les agrégats de dépenses ne prennent pas en compte les différences majeures de mise en œuvre des politiques publiques entre les pays et leur mode de financement. Certaines politiques passent par d'autres canaux que la dépense sociale (dépenses fiscales notamment), dépendent davantage du privé ou reposent sur d'autres postes de dépenses publiques. De Deken (2013) souligne par exemple que la politique de lutte contre la pauvreté en Angleterre et aux États-Unis passe principalement par l'aide au logement (*Housing First*) ce qui n'est pas considéré dans la base de données sur les dépenses sociales de l'OCDE largement utilisée dans les travaux précédemment cités.

Enfin, l'approche par les dépenses ne tient pas compte de la qualité des prestations fournies ce qui limite considérablement sa capacité à mesurer l'effet de ces politiques. La mise en œuvre des politiques d'investissement social peut, à dépense équivalente, prendre des formes très différentes et impacter les bénéficiaires tout aussi différemment (Elbaum, 2010). En outre, certaines dépenses d'investissement social peuvent avoir des effets contradictoires ou indéterminés ; par exemple une allocation parentale qui vise à financer un mode de garde peut inciter les mères à moins ou ne plus travailler si l'élasticité de l'offre de travail est très sensible aux revenus compensatoires, ce qui est souvent le cas pour les bas salaires, accentuant l'effet Mathieu de l'investissement social.

Les analyses qui cherchent à mesurer l'efficacité de ces dépenses se voient opposer d'autres arguments plus méthodologiques. Le problème lorsqu'on cherche à lier les efforts d'investissement social et les résultats observés est d'identifier une relation robuste qui peut s'interpréter comme un lien de cause à effet. Dans la majorité des cas, ce n'est pas possible (Kvist, 2016). Voici quelques arguments généraux expliquant pourquoi.

a) L'endogénéité

Les variables explicatives utilisées dans les analyses sont elles-mêmes endogènes, les interprétations causales des corrélations sont alors incorrectes. Des différences existent entre les pays à la fois sur leurs dépenses et sur leurs performances et ces deux éléments sont effectivement souvent corrélés. Toutefois, la générosité d'un système social peut être à la fois une cause et une conséquence des performances. Le sens de la relation est bien sûr crucial pour émettre des recommandations économiques et les modélisations employées parviennent rarement à séparer les effets croisés à partir de données agrégées. Les dépenses capturent les intrants de la politique (la dépense de l'investissement social) mais aussi une partie de ses effets : l'argent dépensé en raison de l'inefficacité des politiques. Les méthodes pour séparer les effets demandent des hypothèses souvent peu crédibles à l'échelle macroéconomique et des données souvent indisponibles.

b) L'incertitude sur la construction des modèles

Parmi les articles cités, beaucoup cherchent à identifier des relations statistiques entre des variables (dépenses sociales et performances macroéconomiques) dont la forme et le sens de la relation n'est *a priori* pas connu. Les modèles estimés dans ces analyses ne reposent pas, en général, sur une base théorique formalisée qui guiderait les auteurs dans le choix des variables à intégrer. Au contraire, les auteurs cherchent plutôt à démontrer que les modèles macroéconomiques standards sont insuffisants en établissant qu'ajouter des mesures de l'investissement social améliore les prédictions. Sans modèle sous-jacent, il n'y a pas de moyen d'identifier en amont l'ensemble des variables appropriées à intégrer. Par conséquent, les choix des variables explicatives sont basés sur des choix *ad hoc* potentiellement stratégiques ce qui fait également planer le risque d'omettre des variables importantes³.

c) La substituabilité

Les performances et caractéristiques des différents pays ne sont pas la réalisation de modèles communs générant soit la croissance, l'emploi, la pauvreté... et si c'était le cas, il n'y a aucune raison pour que ce modèle soit linéaire comme l'imposent les modèles de régression. Les différences observées ne peuvent se résumer à la contribution de variables de contrôle (souvent endogènes) ou la réalisation de chocs aléatoires dans le processus générant les données. Des lors, les pays ne sont pas interchangeables, ils ne

³ On parle de biais de variable omise dont la représentation algébrique est à trouver chez Wooldridge (2009).

représentent pas ce qu'il se serait passé pour les autres s'ils avaient eu les mêmes caractéristiques.

Ces critiques ne sont pas spécifiques à l'investissement social. On retrouve des problèmes similaires dans les travaux qui analysent les déterminants de la croissance (Durlauf, 2009) dont nous avons d'ailleurs soulevé les enjeux dans la note France Stratégie, *L'éducation peut-elle favoriser la croissance ?* (Heim & Ni, 2016). Ces limites sont bien connues des auteurs (voir notamment la conclusion de De Deken (2016)) qui prônent généralement la combinaison d'approches macro avec des analyses micro ou méso, plus robustes mais difficilement généralisables.

Cette courte discussion des recherches existantes permet de dresser plusieurs constats. D'abord, les comparaisons internationales des performances macroéconomiques des pays en fonction de leur niveau d'investissement social interrogent l'efficacité de l'investissement social. Elles mettent en exergue des décalages entre dépenses et résultats mais ne permettent pas d'identifier l'effet de la dépense sociale sur les performances économiques, ni son rendement. Les liens statistiques entre dépenses d'investissement social et performances macroéconomiques n'identifient pas l'impact, c'est-à-dire l'effet causal de l'investissement social. En réalité, les possibilités d'identifier l'impact d'une politique publique au sens causal se limitent à des conditions de mise en œuvre spécifiques et/ou à l'utilisation de techniques économétriques particulières assez peu adaptées aux analyses macroéconomiques.

Dès lors, le niveau d'observation adapté n'est plus macro mais méso : chaque politique sociale voire chaque dispositif d'une politique sociale. L'objectif de l'investissement social devient alors, pour une politique donnée, de maximiser son impact sur la société.

1.2. Quelles preuves pour l'investissement social ?

Après l'exposé de ces travaux, de leurs méthodes, de leurs conclusions et des critiques que nous leur adressons, il apparaît clair que l'évaluation de l'investissement social vise à répondre à plusieurs questions. Les chercheurs et institutions se placent à des niveaux d'observation très différents et font appel à des données et sources d'informations très diverses. La pluralité des méthodes et les doutes qui planent sur la robustesse de certaines créent une certaine confusion dans la littérature où tous les moyens semblent bons pour soutenir ou critiquer l'investissement social. Ce décalage de niveaux de preuve appelle une clarification des enjeux et une hiérarchisation des « *evidences* » qui permettrait d'asseoir un niveau d'exigence et de robustesse.

Encore une fois, ce besoin n'est pas spécifique à l'IS mais il se fait particulièrement ressentir tant la confusion est grande. Pour hiérarchiser les niveaux de preuves, plusieurs classifications existent dans la littérature : Marston et Watts (2003) ont classé

les preuves dans les recherches en sciences sociales comme étant « dures » ou « molles », suggérant implicitement des niveaux objectifs et subjectifs. Les preuves dures sont principalement composées de données quantitatives, les molles sont les données qualitatives telles que les ethnographies, etc.

D'autres auteurs proposent des classifications plus ordonnées (Bagshaw & Bellomo (2008) à gauche, Petticrew & Roberts (2003) à droite :

- | | |
|--|---|
| 1. Évaluations randomisées bien menées et disposant d'une puissance statistique suffisante | I. Revues systématiques et méta-analyses |
| 2. Évaluations randomisées bien conduites mais de petites tailles | II. Évaluations randomisées (RCT) avec résultats reproduits |
| 3. Études observationnelles non randomisées | III. Évaluations randomisées sans résultats consolidés |
| 4. Évaluations non randomisées avec contrôle des caractéristiques passées | IV. Études longitudinales |
| 5. Études de cas sans contrôle | V. Études de cas avec groupe de contrôle |
| | VI. Analyses en coupe |
| | VII. Études de cas sans groupe de contrôle |

Si des divergences existent, ces deux classifications ont au moins un point commun : elles accordent beaucoup de crédit aux expérimentations aléatoires. Un débat semblable à celui sur les classifications des dépenses d'investissement social existe dans le champ classifications des niveaux de preuves : la frontière entre les différents niveaux est fine, masque des complémentarités possibles tout en écartant systématiquement certaines approches. Pourtant, certaines analyses observationnelles et qualitatives peuvent fournir des résultats intéressants et robustes⁴, permettent de répondre à d'autres questions que ce qui marche ou pas, etc. Ces classifications ont aussi le défaut de se focaliser uniquement sur la validité interne des études et s'éloignent de la définition plus générale de la preuve (Rychetnik, Fromme, P, & shiell, 2002).

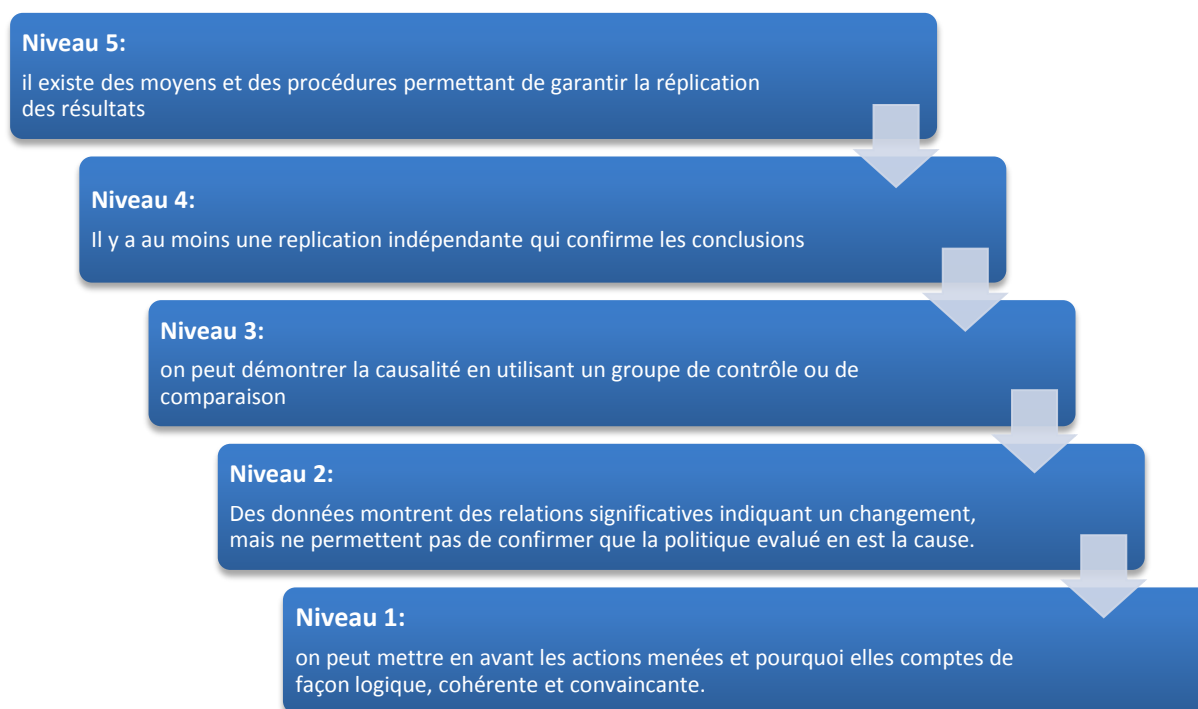
Des organisations produisant des évaluations et des synthèses de littérature sur l'efficacité de politiques sociales ont adoptées une classification explicite de la littérature qu'ils considèrent. Par exemple, celle du *Washington State Institute for Public Policy* (2016) utilisée pour produire des analyses coûts-bénéfices de programmes sociaux est assez proche de celle de Bagshaw & Bellomo (2008) et n'intègrent que les travaux contrôlant a minima les différences dans un modèle de régression avec variable de contrôles.

La hiérarchie proposée par NESTA (2016) adopte un point de vue moins méthodologique que les précédentes, ce qui nous apparaît plus adéquat. Cependant, ces hiérarchisations de preuves masquent les synergies possibles entre différentes

⁴ Benson et Hartz (2000) comparent les résultats observationnels et ceux issus d'évaluations aléatoires contrôlées d'études médicales publiées depuis 1984. Sur les 19 traitements évalués dans 136 évaluations, seuls 2 résultats sont hors de l'intervalle de confiance à 95 %. Néanmoins, les résultats de travaux médicaux sont en général beaucoup moins sujets au biais de sélection que les interventions sociales.

méthodes de recherche et en particulier la complémentarité entre expérimentations et analyse qualitatives. En réalité, les méthodes et les données utilisées ne permettent pas de répondre aux mêmes questions, c'est pourquoi il est préférable de substituer, à la hiérarchie, des typologies. NESTA (2016) en a notamment proposé une discutant les « pour » et les « contre » des différentes méthodes d'évaluation dont nous pouvons nous inspirer pour l'investissement social.

Figure 1 – Les niveaux de preuves tels que défini par le *think tank* NESTA (Alliance for useful evidence, 2016)



Le choix des méthodes appropriées aux questions de recherche auxquelles on souhaite répondre apparaît plus flexible et pragmatique qu'une approche centrée sur les méthodologies. Cependant, on ne doit pas pour autant en conclure que tout est bon pour faire preuve mais que les différentes méthodes ne peuvent pas répondre aux mêmes questions.

L'évaluation de l'impact de l'investissement social peut ainsi être obtenue à partir d'expérimentation, de quasi expérimentations, dans une moindre mesure à partir d'analyses longitudinales. La réplication des résultats dans des contextes différents par des équipes indépendantes est un gage de qualité qui renforce le niveau de preuve. Les revues systématiques apportent alors la plus grande certitude quant à l'efficacité d'une politique d'investissement social.

L'évaluation du rendement de l'investissement social est une étape supplémentaire, *ad hoc*, qui utilise les résultats des évaluations d'impact où de méta-analyses, des valorisations de ces effets en termes monétaires et des coûts associés.

Notre analyse ne concerne que ces deux questions qui font appel à des approches microéconomiques. Nous ne développons pas en détail les enjeux d'évaluation *ex-ante*. Nous discutons certains points les concernant dans nos recommandations.

1.3. Évaluation d'impact et investissement social

Les évaluations micro de l'évaluation des politiques d'investissement social sont caractérisées par les données utilisées et l'emploi de méthodes d'évaluation d'impact dont nous présentons brièvement les principaux enjeux. Le récent rapport de Desplatz et Ferracci (2016) discute plus en détails les différentes méthodes et leurs implications⁵. Nous ne faisons que rappeler des éléments d'intuition et des exemples emblématiques pour les illustrer.

1.3.1. Les difficultés de l'évaluation d'impact

Identifier l'impact des politiques d'investissement social suppose de comparer la réalité avec celle qui prévaudrait si la politique n'avait pas été mise en place. Il s'agit alors d'estimer cette réalité alternative pour pouvoir comparer les situations.

L'identification d'un **contrefactuel** plausible est la principale difficulté. En théorie, l'impact est la différence entre l'indicateur de résultat pertinent en présence du programme et sans. Alors que la situation après intervention est observée et mesurée, celle en l'absence de programme ne l'est pas. Le problème de l'évaluation est en fait un problème de données manquantes qui nécessite de se servir d'un groupe de comparaison pour représenter le contrefactuel. Les deux sources de comparaisons naturelles qui existent : la comparaison avant/après et la comparaison des populations affectées/non affectées ne suffisent généralement pas à estimer de façon satisfaisante ce qui se serait passé.

Dans le premier cas, l'évolution observée entre les différentes dates peut être due en partie à la politique évaluée, mais aussi à l'effet d'autres politiques, à une évolution naturelle tendancielle, etc. Bref, à une multitude de facteurs, souvent inobservables, dont l'effet est confondu et indissociable (en général) de celui de la politique dont on cherche à mesurer l'impact.

Dans le second cas, les populations affectées et non affectées ont de grandes chances d'être très différentes. Même si au sein d'une population d'éligibles, tout le monde ne bénéficie pas de la politique, leur comparaison revient à estimer ce qui est dû à la politique mais aussi à l'effet de leurs différences et des raisons pour lesquels ils n'en bénéficient pas.

Il peut être envisagé de raisonner *toutes choses égales par ailleurs* et de contrôler (lorsque les données sont disponibles) pour un ensemble de différences et « neutraliser » leur effet. L'objectif est alors de rendre virtuellement identique les

⁵ Desplatz R. et Ferracci M. (2016), [*Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens*](#), Paris, France Stratégie.

individus sur la base de leurs caractéristiques observables pour pouvoir les comparer. Dans certains cas, cette démarche peut suffire à identifier un effet qui reflète assez bien la relation causale, à partir de méthodes économétriques simples (régressions, matching, etc.). Cependant, il existe souvent d'autres caractéristiques inobservables pour le chercheur qui conditionnent à la fois l'accès à la politique sociale et les *outcomes*. Elles créent alors un biais de variable omise qui affecte l'identification de l'effet de la politique publique.

Ainsi, la description de l'évolution de la situation depuis la mise en œuvre des dispositifs ne constitue pas une mesure de l'impact ; le constat que des objectifs ont été remplis ne garantit pas que ce soit le résultat de la politique mise en œuvre ; la comparaison entre les bénéficiaires et les non-bénéficiaires d'une politique publique donne rarement une bonne mesure de l'effet de cette dernière (voir Annexe).

Les solutions des méthodes d'évaluation d'impact consistent alors à exploiter des événements plausiblement aléatoires qui modifient les chances d'être affecté par une politique d'investissement social.

a) L'expérimentation aléatoire pour l'évaluation d'impact

Ces changements aléatoires peuvent être directement provoqués dans des expérimentations aléatoires contrôlées, qui constituent l'étalon d'or de l'évaluation d'impact. En assignant directement au hasard les bénéficiaires de la politique, on s'assure qu'il n'y a pas, en moyenne, de différence entre les bénéficiaires et le groupe de contrôle et qu'alors, le seul écart observé est dû à la politique. Plus spécifiquement, le tirage aléatoire élimine le biais de sélection/d'échantillonnage, contrôle pour toutes les variables même inobservables, ce qui améliore grandement les chances d'identifier une relation causale. L'expérimentation aléatoire contrôlée fournit une estimation rigoureuse et objective de l'effet de la politique publique. Elle n'est pas dénuée de défaut pour autant. Les principales critiques sont exprimées dans Heckman (1992) pour qui l'expérimentation est une analyse sans théorie⁶. Elle ne vise pas à estimer de paramètres structurels mais compare simplement des résultats moyens. Il manque alors, pour Heckman, les clés pour comprendre et pour expliquer les phénomènes observés et surtout pouvoir inférer de ces résultats des conclusions plus générales. Ensuite, l'approche expérimentale suppose que le protocole ne modifie pas la situation observée (absence d'effet Hawthorne⁷ ou John Henry⁸), ce qui est une hypothèse parfois forte et difficilement testable. Enfin, l'expérimentation ne permet pas d'obtenir d'autres

⁶ Il y a bien sûr des fondements théoriques statistiques mais les expérimentations sociales reposent rarement sur un modèle structurel dont il s'agit d'estimer les paramètres. Les sophistications statistiques dans les recherches expérimentales ont généralement pour objectif d'asseoir la robustesse des estimations ou de résoudre des problèmes d'attrition, de non-réponse, de contamination entre les groupes, etc.

⁷ Lorsque les résultats d'une expérience ne sont pas dus aux facteurs expérimentaux mais au fait que les sujets ont conscience de participer à une expérience dans laquelle ils sont testés, ce qui se traduit généralement par une plus grande motivation.

⁸ Dans une expérimentation, si le groupe de contrôle a conscience de son statut « non traité », il peut réagir activement pour compenser le désavantage d'être dans le groupe de contrôle.

paramètres que l'effet moyen⁹ (l'effet médian par exemple ne peut s'obtenir qu'au prix de nouvelles hypothèses diminuant l'intérêt de la démarche expérimentale¹⁰). Nous ajoutons qu'il est souvent, en pratique, très compliqué de constituer un groupe de contrôle, c'est une démarche coûteuse et assez mécanique, qui ne renseigne pas seul sur les raisons de l'effet (lire L'horty & Petit (2010) ou Banerjee & Duflo (2009) pour une discussion).

Les expérimentations aléatoires contrôlées sont souvent critiquées pour leur caractère *local*¹¹ et leur supposée incapacité à être appliquées à des projets de grande ampleur ou à identifier les effets généraux. Or, les décideurs sont souvent demandeurs d'impacts systémiques. Ces résultats s'estiment difficilement et rarement à partir d'évaluations d'impact expérimentales. Les outils d'équilibre général calculable et de microsimulations sont plus adaptés mais leurs conclusions sont limitées et reposent sur des hypothèses fortes. Pour autant, les évaluations expérimentales ne sont pas que des réponses partielles sur l'efficacité de micro-dispositifs. Des évaluations de grande ampleur ont été mises en place et ont permis de répondre à des questions de politique publique importantes, y compris dans le champ de l'investissement social. Dans le domaine de l'éducation par exemple, Fryer (2016) recense 196 expérimentations aléatoires mises en place dans les pays développés¹². De nombreuses expérimentations ont également été menées pour inciter au retour à l'emploi – en France, voir notamment Parent, Sautory et Desplatz (2013) pour une synthèse des résultats. Les évaluations randomisées peuvent également avoir des designs complexes et tester des interactions entre différentes politiques ou des comparaisons d'interventions. Un exemple est à trouver dans le travail de Duflo, Dupas et Kremer (2011) qui ont testé à la fois une variation de la taille des classes au primaire au Kenya, une formation des équipes de direction sur l'encadrement des nouveaux professeurs recrutés et une individualisation de la pédagogie en fonction des résultats passé des élèves.

b) Les aléas naturels : l'utilisation de méthodes quasi expérimentales.

Une autre façon de produire des évaluations d'impact est d'identifier ce qui, dans la nature ou les règles institutionnelles, modifie un peu au hasard la probabilité de bénéficier d'une politique d'investissement social. Par exemple, une règle qui oblige un enfant à avoir deux ans au 1^{er} Janvier pour être accepté en maternelle est assez arbitraire. Le jour de naissance peut être considéré comme aléatoire (au moins localement autour du 1^{er} janvier (Grenet, 2010)) pour qu'en moyenne, la seule différence entre un enfant né un 1^{er} janvier N-2 et un 31 décembre N-3, soit l'accès à l'école maternelle à 2 ans plutôt qu'à 3. C'est un exemple d'expérience naturelle qui a été

⁹ L'inférence statistique de l'effet du traitement par quantile fait appel à des hypothèses d'identification fortes que la randomisation ne suffit pas à valider et pose des problèmes d'interprétation importants (Imbens & Wooldridge, 2008).

¹⁰ « Experimental methods cannot estimate median differences without invoking stronger assumptions than are required to recover means ». (Heckman J. J., 1992).

¹¹ À la fois dans leur dimension non généralisable et dépendante du contexte.

¹² Parmi une sélection de près de 800 travaux parmi lesquels près de 400 ont été exclus car ils n'utilisaient pas de mesure de performance en mathématiques ou en littérature.

exploité par Goux et Maurin (2010) pour tester l'effet de la disponibilité de place en maternelle sur l'emploi des mères (voir *infra*).

Certaines politiques publiques conditionnent leur accès au respect de règles, parfois arbitraires, qui séparent la population en un groupe d'éligibles et un groupe de non-éligibles. La comparaison de ces deux groupes ne fournirait pas une bonne estimation de l'effet de la politique étudiée. Cependant, si la règle est nette, non manipulable et qu'elle est basée sur une variable continue (autour d'un seuil de revenu par exemple), on peut s'intéresser plus en détail aux individus qui sont tout juste éligibles et tout juste non-éligibles. En effet, les individus autour de ce seuil ont de fortes chances de beaucoup se ressembler si bien qu'à quelques détails près certains vont pouvoir bénéficier d'une politique publique alors que d'autres non. En moyenne, tout se passe comme si, autour de ce seuil, des personnes très semblables avaient été tirées au hasard pour être éligibles à la politique et d'autres non. La discontinuité de régression est assimilable à une expérimentation aléatoire locale (Lee & Lemieux, 2010). Fack & Grenet (2015) utilisent la discontinuité d'accès aux bourses sur critères sociaux en fonction du revenu des parents pour tester leur impact sur la poursuite d'études supérieures (voir *infra*).

Les règles d'attribution des politiques ne sont pas les seuls éléments qui génèrent des situations quasi aléatoires. La *nature* crée parfois des situations qui modifient les probabilités d'être affecté par une politique publique. Ces situations sont toutefois beaucoup plus rares et reposent sur des hypothèses parfois difficiles à défendre.

On peut citer quelques travaux emblématiques utilisant ce type de stratégie d'identification : Angrist & Krueger (1991) combinent les différences de date de naissance et de règles d'âge d'entrée à l'école qui varient entre les États américains pour évaluer le rendement de l'éducation. Card (1995) remarque que plus la distance entre le lieu de résidence et l'université est élevée, moins les jeunes poursuivent d'études supérieures. Il exploite alors la distance à une université pour identifier l'effet d'avoir un diplôme du supérieur sur les salaires. Angrist et Evans (1998) constatent que les parents ont une préférence pour la mixité pour leurs enfants et qu'alors, les couples dont les deux premiers enfants sont de même sexe ont plus de chance d'avoir un troisième enfant que les couples dont les deux premiers enfants sont de sexe opposé. Ils se servent donc de l'aléa du sexe des deux premiers enfants pour estimer l'effet d'avoir un troisième enfant sur l'emploi des femmes.

La principale limite de cette méthode tient à la rareté des instruments disponibles et aux hypothèses d'application. Il n'y a que rarement des variables qui influencent fortement le traitement sans affecter le résultat et, *a contrario*, les variables vraiment exogènes ont souvent une influence faible sur le traitement, ce qui limite les possibilités d'inférence. De plus, les instruments sont parfois assez anecdotiques donc peu reproductibles, ce qui peut faire douter de la validité externe des résultats qui en découlent.

Les politiques d'investissement peuvent aussi être déployées de façon progressive et créer ainsi des groupes de comparaison qui peuvent servir à identifier l'effet de la

politique dans certaines conditions. Il est alors nécessaire d'avoir des données avant et après la mise en œuvre de la politique dans les zones¹³ affectées et non affectées (ou pas encore) et que les zones test et contrôle se ressemblent suffisamment pour qu'il soit plausible que leurs destins aient été les mêmes en l'absence de politique¹⁴. L'identification de l'effet tient alors à la double différence avant-après entre les zones affectées-non-affectées. C'est notamment le cas de la politique anglaise *Pathway to work* (voir *infra*), de l'expérience de transferts conditionnels au Mexique *Progressa*, etc.

L'utilisation de ces méthodes quasi expérimentales doit toutefois être abordée avec une extrême précaution car elles sont moins transparentes et plus sujettes à des divergences d'opinions quant à leur validité (Duflo, 2005). La communication de leurs résultats peut également être compliquée à cause de toutes les mises en garde qui doivent les accompagner qui peuvent amener de la confusion ou induire en erreur lorsqu'elles ne sont pas évoquées.

Si les évaluations d'impact sont difficiles à mettre en place et parfois à interpréter ou à extrapoler à des populations plus larges, elles sont aujourd'hui les seules méthodes permettant d'interpréter une relation entre une intervention et un résultat comme un lien de cause à effet avec un minimum d'hypothèse. Cette validité interne ne garantit pas que ces résultats seront obtenus pour d'autres publics, dans d'autres conditions ou à un autre moment. Les chances que ce soit le cas dépendent de la qualité de l'échantillon à partir duquel les conclusions sont tirées (un échantillon aléatoire, sans trop d'attrition améliore la validité externe) et du contexte (plus les conditions de l'évaluation sont proches de la réalité une fois généralisée, meilleures sont les chances de retrouver un effet). Si l'expérience est renouvelée ailleurs, le fait que le résultat soit retrouvé dans d'autres contextes est également un gage de qualité.

C'est donc la multiplication des évaluations de qualité d'investissement sociaux semblables dans leur mise en œuvre mais répliquées dans des contextes différents et la synthèse de ces effets mesurés qui permettront de limiter les risques d'échec des politiques.

1.3.2. Quelques évaluations d'impact de politiques d'investissement social

Le but de cette section est de montrer qu'il existe un nombre important d'évaluations d'impact rigoureuses de politiques d'investissement social, y compris en France et qu'un travail de revue systématique et critique est un impératif pour pouvoir évaluer l'efficacité de l'IS et ensuite estimer son rendement. Par souci de synthèse nous nous focalisons

¹³ Ou groupe d'individus. Le déploiement géographique progressif est souvent plus propice car plus acceptable d'un point de vue éthique.

¹⁴ L'idéal étant de tirer au hasard les localités ou groupes qui profitent en premier ou plus tard du programme.

sur trois grandes thématiques de l'investissement social : la petite enfance, l'éducation et les politiques d'activation. Le propos de cette note n'est pas de recenser de façon exhaustive cette littérature. Les conclusions présentées sont d'ailleurs souvent différentes d'une étude à une autre. Nous ne nous attachons pas à décrire toutes les méthodologies employées pour ne pas trop alourdir cette analyse déjà dense. Nous ne signalons que quelques éléments (et notamment s'il s'agit d'expérimentation aléatoire). En revanche, puisque ce travail n'est pas une revue systématique, nous discutons des interprétations et des mécanismes expliquant ces résultats plus que de leur magnitude.

Les dispositifs évalués vont de petites expérimentations (aléatoires ou non) à des politiques plus ambitieuses et générales. Ces études ne parlent généralement pas d'investissement social. Le choix de les intégrer tient à notre compréhension des objectifs de ces politiques et aux résultats qu'elles produisent. Ainsi, les travaux cités ci-après concernent des dispositifs qui visent directement à permettre aux individus de produire, maintenir ou faire bon usage de leur capital humain.

a) Quelques éléments sur les investissements dans la petite enfance.

Les évaluations d'investissement dans la petite enfance font largement références aux projets américains *Perry Preschool Program* ou *Abecedarian Project* qui concernaient un nombre très restreint d'enfants défavorisés aux Etats-Unis. Ces programmes intensifs, expérimentaux ont permis de suivre les enfants plusieurs décennies, de mesurer régulièrement leurs progrès et de constater l'impact de ces programmes d'interventions précoces une fois adulte. Les évaluations de ces projets mettent en avant des effets positifs importants sur le développement cognitif qui disparaissent rapidement (et ne sont plus significatifs à l'âge de 8 ans). Cependant, de nouveaux effets apparaissent à l'âge adulte sur le marché du travail et les conditions de vie sociale (criminalité, comportement à risque) [(Barnett, 1995); (Karoly, Kilburn, & Cannon, 2005) ; (Heckman, S., Pinto, Savelyev, & Yavis, 2010), (Elango, García, Heckman, & Hojman, 2015)]. Les décompositions de ces effets montrent par ailleurs que la diminution des comportements extériorisés (comportement social négatif incluant notamment agression et délinquance,...) explique une part importante des effets observés sur les adultes (Heckman, Pinto, & Savelyev, 2013).

Cependant, les évaluations de politiques visant à élargir l'accès à des modes de garde n'ont pas toujours permis de montrer un effet positif. Par exemple, Baker & Milligan (2008) au Canada ou Datta Gupta & Simonsen (2010) au Danemark montrent que plus les enfants passent de temps en crèche, moins leurs compétences, cognitives et conatives se développent, alors que Havnes & Mogstad (2011) exploitent les différences d'expansion d'établissements subventionnés d'accueil des jeunes enfants en Norvège et concluent à un effet positif et durable de l'accès à un mode de garde sur les performances scolaires et, beaucoup plus tard, sur l'emploi des enfants ayant bénéficié de ces modes de garde.

Ces divergences apparentes peuvent être expliquées de trois façons (Blanden, Bono, McNally, & Rabe, 2015) :

- la durée de la garde proposée ;
- la qualité du mode de garde ;
- la qualité des alternatives en l'absence d'intervention.

L'effet de l'intensité est difficile à anticiper car d'un côté les enfants peuvent manquer d'interactions avec leurs parents, ce qui peut avoir des effets négatifs sur leur développement, de l'autre, il améliore l'employabilité des parents et peut améliorer les conditions économiques de la famille. La qualité du mode de garde est à comparer avec la situation contrefactuelle. Si le mode de garde proposé est de grande qualité et que l'alternative en l'absence de la politique est de mauvaise qualité, alors on attend des effets positifs. Bernal & Keane (2011) exploitent des changements de législation entre États américains conduisant les mères célibataires à avoir davantage recours à un mode de garde pour pouvoir travailler et évaluent alors l'effet d'être gardé sur les performances scolaires des enfants. Ils montrent alors qu'en moyenne, une année d'utilisation de mode de garde réduit les résultats scolaires des enfants de 0.1 écart-type. Néanmoins ce résultat masque des différences importantes suivant les modes de garde. En effet, ceux formels et basés en centre n'ont pas d'effet sur les performances des enfants. Seuls les modes de gardes informels – grands-parents, familles ou autres – ont un impact négatif significatif.

De façon similaire, les premières conclusions de Garcia, Hojman & Shea (2016)¹⁵ – évaluant un dispositif expérimental offrant l'accès gratuit à un programme d'éducation pour les tout-petits – suggèrent une substituabilité entre temps passé avec les parents et dans les structures de garde où le capital humain des parents est déterminant. Les enfants défavorisés – dont le capital humain des parents est limité – bénéficient alors davantage du programme éducatif que les enfants plus favorisés, pour qui la garde formelle se substitue simplement à la garde parentale, ce qui, en moyenne, ne change pas grand-chose pour ces enfants.

Au début des années 2000, l'Angleterre a mis en place une politique d'accueil gratuite des enfants de 3 ans qui s'est déployée progressivement sur le territoire. Ces nouvelles places ont été ouvertes dans des structures privées mais avec des financements publics. L'évaluation de l'effet de cette politique proposée par Blanden, Bono, McNally, & Rabe (2015) montre un faible effet positif à court terme sur les performances des enfants qui disparaît très vite, ce qu'ils expliquent par le manque d'intensité du recours, et le peu de gain de qualité entre les nouveaux modes de garde et la situation contrefactuelle. En réalité, pour 10 places financées, seules 2 ont effectivement été créées, ce qui indique un effet d'éviction des modes de garde alternatifs plus qu'une augmentation de l'enrôlement préscolaire. C'est un résultat qu'on retrouve également en France (Maurin &

¹⁵ Ce travail est une version préliminaire d'un article que les auteurs nous ont directement transmis et nous les en remercions.

Roy, 2008). La qualité des modes de garde ainsi proposés est également fortement questionnée par les auteurs, qui corroborent leur doute par le constat que dans ces nouveaux centres d'accueil des enfants, moins d'un employé sur cinq est diplômé alors qu'ils le sont quasiment tous dans le public. La qualification des intervenants est pourtant identifiée comme l'une des caractéristiques cruciale qui détermine la qualité de l'accueil des enfants en bas âge (Ackerman, 2005) (Florin A. , 2007). Une revue systématique de littérature financée par *Campbell Collaboration* est en cours de production dont le protocole de recherche est disponible (Manning, Garvis, Fleming, & Wong, 2015).

Dans leur revue de la littérature sur l'efficacité des interventions précoces (non américaines), Nores et Barnett (2010) rapportent des effets positifs, significatifs, importants et durables, en particulier lorsque les interventions intègrent une composante de stimulation cognitive. La majorité des interventions précoces efficaces ont débuté entre 3 et 6 ans (Almond & Currie, 2011) mais, le gradient socioéconomique des compétences cognitives, des fonctions socio-émotionnelles et l'état de santé, existe déjà à 3 ans, ce qui suggère que des programmes agissant plus tôt pourraient être encore plus efficaces.

D'un autre côté, une part importante de la littérature traitant de l'âge d'entrée à l'école tend à montrer que commencer l'école plus tard améliore les performances scolaires (Datar, 2006); (McEwan & Shapiro, 2008) ainsi que le développement non-cognitif (Mühlenweg, Blomeyer, Stichnotha, & Laucht, 2012). D'autres recherches n'identifient pas d'effet (Fertig & Kluge, 2005) ou des effets positifs d'une entrée précoce (Suziedelyte & Zhu, 2015). Ces travaux utilisent des stratégies d'identification très proches exploitant des règles d'âge d'entrée à l'école ou d'âge théorique comme variable instrumentale ou discontinuité de régression. Cette apparente contradiction interroge la nature des interventions précoces pour laquelle le choix de la scolarisation n'est pas forcément optimal.

Les recherches sur l'effet d'une scolarisation précoce sur les performances des élèves sont assez rares en France. Des résultats statistiques suggèrent que les enfants qui entrent en maternelle à deux ans plutôt qu'à 3 ont de meilleurs résultats scolaires au primaire et entrent plus tôt à l'école élémentaire (Caille, 2001). Cependant, il n'est pas certain que ce qui est mesuré corresponde à l'effet causal de la maternelle à deux ans ou à des différences préexistantes que le statisticien ne peut intégrer dans son analyse. L'étude similaire menée par Caille & Rosenwald (2006) ne permet d'ailleurs pas de conclure à un effet positif de l'entrée en maternelle à 2 ans plutôt qu'à 3.

À notre connaissance, en France, seules trois études identifient des « expériences naturelles » d'accès à la maternelle à deux ans pour estimer l'effet sur les performances scolaires.

Filatriau, Fougère, & Tô (2013) exploitent les contraintes locales de places disponibles pour accueillir les enfants de 2 ans en maternelle pour identifier l'impact d'une scolarisation précoce et trouve un effet positif sur le développement cognitif et non cognitif à 6 ans, puis sur les performances en mathématiques et en français en 6^e et 3^e.

Dumas et Lefranc (2010) emploient également une stratégie de variable instrumentale pour estimer l'effet de l'âge d'entrée en maternelle sur les résultats. Ils utilisent pour cela l'âge moyen d'entrée en maternelle dans le département pour une cohorte donnée et identifient des impacts négatifs sur la probabilité d'avoir redoublé à 16 ans et un effet positif sur la probabilité d'avoir le bac. On note toutefois que cette stratégie est fragile car l'instrument utilisé est susceptible d'affecter les performances scolaires par d'autres canaux que l'âge d'entrée en maternelle, ce qui invaliderait l'estimation. Si la maternelle a des effets positifs sur les performances, les régions où les enfants entrent plus tôt sont en moyenne composées de meilleurs élèves, ce qui peut améliorer les performances des élèves qui bénéficient de pairs de meilleurs niveaux.

Ces études (et celles étrangères précédemment citées) confondent l'impact de l'âge d'entrée avec trois autres effets : la durée de la scolarité, l'âge de l'enfant au test et son âge relatif dans sa classe. Les deux premiers sont parfaitement colinéaires avec l'âge d'entrée à l'école. Le troisième est fortement corrélé. Séparer ces effets implique de disposer d'instruments crédibles ou d'imposer des contraintes fortes sur la forme des relations. À ce titre, les travaux de Crawford *et al.* (2013) utilisent différentes sources de données et techniques économétriques pour séparer ces effets. Ils montrent alors que l'âge au test est le facteur le plus responsable des différences de performances selon le mois de naissance. L'effet propre de l'âge d'entrée serait quant à lui nul d'après leurs estimations.

Dans leur travail focalisé sur l'impact sur l'emploi des femmes, Goux et Maurin (2010) s'interrogent également sur l'effet de la scolarisation à deux ans sur la probabilité d'arrêter l'école entre 16 et 18 ans. Constatant que les différences entre les régions à fort ou faible taux de scolarisation à deux ans sont beaucoup plus marquées pour les enfants nés pendant les mois d'été, ils se servent de ces variations de probabilité d'accès à la maternelle pour en identifier l'effet. Ils ne trouvent pas de relation significative entre l'accès en maternelle à 2 ans et les sorties précoces du système scolaire. Cette mesure de la réussite éducative est toutefois trop imprécise pour distinguer aisément un effet à partir de variations régionales de taux de scolarisation à deux ans.

En revanche, leur travail sur l'emploi des mères exploite la discontinuité d'accès à l'école maternelle à deux ans pour les enfants nés entre décembre et janvier. Ils montrent alors que l'accès à une place en maternelle à deux ans augmente la participation au marché du travail des mères célibataires, mais pas celle des femmes en couple.

Goux et Maurin ne sont pas les seuls auteurs à montrer que l'investissement dans les modes de garde ou la scolarité précoce permet aussi de stimuler l'emploi des femmes. Une littérature internationale abondante et active produit des résultats contrastés en exploitant généralement des différences locales d'accès aux crèches (Bauernschuster & Schlotter, 2015) ou des règles d'âge des enfants (Graves, 2013). D'autres s'intéressent aux coûts des modes de garde et aux politiques de financement pour mesurer la réaction des femmes. (Givord & Marbot, 2015) montrent ainsi que la mise en place de la

PAJE en France a eu un impact positif significatif sur le recours à des modes de garde payants et se traduit par un effet positif sur l'emploi des femmes, particulièrement celles qui ont plusieurs enfants. Ce résultat semble indiquer que la PAJE corrige en partie l'effet désincitatif à l'emploi de dispositifs passés comme l'allocation parentale d'éducation¹⁶.

Les travaux présentés ici indiquent que l'investissement dans la petite enfance peut être bénéfique aux enfants mais une revue plus systématique et différenciée, suivant le type d'investissement et l'âge de l'enfant, est nécessaire pour conclure à la fois sur l'efficacité et sur les conditions pour que ce soit le cas.

b) Des investissements dans l'éducation

L'investissement dans l'éducation est souvent appréhendé par la dépense par élève et le ratio élèves/enseignants. L'évaluation de l'efficacité d'une augmentation des ressources éducatives (ou du ratio enseignant/élève) n'est pas simple. Les ressources peuvent être dirigées vers une partie d'élèves en difficulté (l'éducation prioritaire) ou au contraire à un groupe favorisé (les classes préparatoires aux grandes écoles). Ainsi, lier les performances des élèves aux ressources éducatives qui leur sont dédiées ne mesure pas leur impact. Une expérience unique a été mise en œuvre pour mesurer l'effet des ressources éducatives sur les performances des élèves. En 1986, dans le Tennessee, 11 600 enfants de dernière année de maternelle à l'équivalent du CE2, parmi 79 écoles ont été assignés au hasard dans des classes de taille normale (± 22 élèves), des classes de petite taille (± 15 élèves) ou des classes de taille normale avec un enseignant supplémentaire. Les enseignants ont également été assignés au hasard. Les résultats de cette expérimentation montrent qu'une augmentation des ressources aux élèves via une taille de classe réduite améliore significativement leurs résultats à la fin de l'année (Krueger, 1999), mais ces effets disparaissent assez rapidement lorsque les élèves retrouvent des classes de taille normale. Néanmoins, (Chetty, *et al.*, 2010) mettent en évidence des effets du programme sur le revenu et la probabilité d'avoir été à l'université. De nombreuses études exploitant des expérimentations naturelles obtiennent des résultats similaires dans différents contextes¹⁷. Il reste que l'effet observé est une forme réduite, comprenant toutes les transformations de l'enseignement qu'une taille de classe plus restreinte permet¹⁸. Investir dans l'éducation pour améliorer le capital humain des jeunes, c'est donc aussi et surtout agir sur la façon d'acquérir des connaissances et de les mobiliser. Pour cela les interventions améliorant les performances des enseignants apparaissent déterminantes.

¹⁶ (Piketty T. , 1998) (Piketty T. , 2003) (Moschion, 2007).

¹⁷ (Angrist & Lavy, 1999) en Israël, (Fredriksson, Öckert, & Oosterbeek, 2013) en Suède, (Browning & Heinesen, 2007) au Danemark, (Urquiola, 2006) en Bolivie, et en France : (Piketty & Valdenaire, 2006) (Bressoux, Kramarz, & Prost, 2009), (Bressoux & Lima, 2011), (Gary-Bobo & Mahjoub, 2013)).

¹⁸ Les élèves ont une attitude « davantage centrée sur la tâche » et les enseignants consacrent moins de temps à la gestion de la classe (Evertson & Randolph, 1989); (Blatchford, Goldstein, Martin, & Browne, 2002). Les nuisances et les effets de saturation dus à une classe trop remplie peuvent également être réduits dans une classe plus petite et jouer un rôle sur les performances (Lazear, 2001).

En France, deux travaux récents ont évalué l'effet de modifications expérimentales des pratiques pédagogiques des enseignants sur les performances en lecture d'élèves de maternelle. Ecalle *et al.* (2015) évaluent une expérimentation où l'enseignant met en pratique des méthodes d'enseignement fondées sur la recherche et trouvent des effets positifs sur certaines dimensions de pré-lecture. Bouguen (2016) évalue une formation d'enseignants encourageant la différenciation pédagogique en fonction du niveau des élèves et obtient également des résultats positifs en reconnaissance des lettres et en phonologies. Chesné (2014) a élaboré une formation des enseignants de mathématiques en 6^e basée sur l'analyse des erreurs des élèves et sur l'élaboration d'une stratégie didactique centrée sur le calcul mental puis évalué l'impact de ces nouvelles pratiques sur les progrès des élèves ; il obtient des résultats très positifs¹⁹.

L'action sur les capacités non cognitives des enfants peuvent également produire des effets importants y compris à long terme. C'est notamment ce que montre l'expérimentation évaluée par Algan, Beasley, Vitaro, & Tremblay (2016) où des enfants défavorisés canadiens de 7 ans ont bénéficié d'un programme expérimental visant à renforcer leurs compétences sociales et leur *self-control*. Après 2 ans d'intervention, les auteurs montrent qu'au début de l'adolescence, le programme limite le risque de comportement agressif, limite l'impulsivité et améliore la confiance. Les effets post-adolescence se manifestent par des performances scolaires améliorées, un moindre comportement agressif et une plus grande confiance. Une fois adulte, les bénéficiaires du programme ont une probabilité plus élevée d'être actifs à temps-plein (en formation ou en emploi), d'être engagés dans un groupe social ou civique, d'avoir obtenu un diplôme du secondaire et d'avoir commis un crime.

Le tutorat et le soutien scolaire sont des formes d'investissements dans l'éducation qui peuvent être fournis de façon publique ou privée, gratuite ou payante. Dans certains pays, le tutorat privé fait partie intégrante de la formation et la dépense des familles atteint le niveau de la dépense publique d'éducation comme en Turquie ou en Corée du Sud (ou elle atteint 2.9 % du PIB (Dang & Rogers, 2008)). En France, le programme de tutorat TALENS – un dispositif parmi les nombreuses « cordées de la réussite » visant à faciliter l'accès aux grandes écoles aux élèves plutôt défavorisés – a été évalué par Riegert, Maurin, & Ly (2016) dans le cadre d'une expérimentation aléatoire contrôlée financée par le Fonds d'expérimentation pour la jeunesse. Profitant d'une demande importante pour participer au programme, les candidats ont accepté d'être tirés au sort pour participer à TALENS et un groupe de contrôle composé de volontaires mais non bénéficiaires a ainsi pu être mis en place. Les jeunes TALENS bénéficient alors d'une semaine de stage à l'École normale supérieure (rue d'Ulm à Paris) en début de première, et de sessions de travail hebdomadaires en petits groupes les samedis après-midi. Les tuteurs sont des élèves normaliens qui interviennent afin de les préparer à l'enseignement supérieur. Les résultats de l'expérimentation indiquent un effet moyen nul sur les notes au baccalauréat, sur l'inscription et la persistance en classe

¹⁹ La méthode d'évaluation a toutefois été critiquée car l'allocation dans les groupes traités et contrôle n'est pas purement faite au hasard.

préparatoire aux grandes écoles. En moyenne donc, TALENS n'améliore pas la réussite scolaire de ses bénéficiaires. Les chercheurs ont procédé à des analyses en sous-groupes et constatent alors que l'effet moyen nul dissimule en réalité un effet positif parmi les meilleurs élèves en seconde, et un effet négatif parmi les élèves volontaires mais moins bons en seconde. En somme, TALENS accentue les inégalités scolaires et dégrade les performances d'élèves moins bons, pourtant volontaires et aussi assidus. Des analyses complémentaires supportent l'idée que ces élèves ont subi l'intensité du programme, trop coûteux en temps. Ainsi, ce programme investissant dans les compétences de jeunes plutôt défavorisés pénalise les plus en difficulté, ce qui questionne la sélection des élèves, le rythme et le contenu du programme ou encore la formation des tuteurs.

Les évaluations peuvent également porter sur des ensembles de dispositifs nationaux plus vastes et sortir d'actions locales, expérimentales ou ciblées.

En Angleterre, le programme *Literacy Hour* a été mis en œuvre dans les écoles primaires à partir de 1998 afin d'améliorer les compétences en lecture des élèves à partir de contenus et pratiques d'enseignement différents. L'évaluation menée par McNally et Machin (2008) met en exergue des effets positifs et significatifs mais d'amplitude modeste. D'après leurs estimations, le fait d'être dans une école participant au programme améliore les performances en pré-lecture des enfants de 2 à 3 points de pourcentage. Cependant, ces effets sont obtenus à partir d'une intervention très peu coûteuse. Le ratio coût/efficacité de ce dispositif apparaît donc très élevé (voir *infra*).

Le programme anglais *Every child matters* qui vise à remédier aux difficultés des élèves ayant des besoins éducatifs spécifiques modérés a également été évalué de façon rigoureuse. Keslair, Maurin, & McNally (2012) montrent que le déploiement décentralisé du programme engendre des variations importantes d'accès pour les élèves entre les localités mais que ces différences ne se retrouvent pas dans les performances scolaires des élèves. Au sein des établissements, les auteurs montrent également qu'il n'y a pas d'effet global du programme sur les élèves.

En France, les programmes de réussite éducative déployés depuis 2005 ont pour objectif d'apporter aux enfants des quartiers prioritaires de la politique de la ville un parcours individualisé reposant sur une approche globale des difficultés rencontrées en mobilisant les acteurs sociaux, médicaux, éducatifs, etc., en collaboration étroite avec la famille. En 2014, ils concernaient environ 100 000 enfants. L'évaluation produite par l'institut des politiques publiques (Bressoux, Gurgand, Guyon, Monnet, & Pernaudet, 2016) met en exergue des effets plutôt négatifs. Le bien-être psychologique des élèves, leur comportement et leurs compétences en mathématiques évoluent de façon similaire à ceux du groupe de contrôle, mais des dimensions capturant la relation aux autres et certaines dimensions de la motivation scolaire sont négativement affectées par le programme. Le seul effet positif est l'assiduité scolaire qui se voit améliorée. Comparé à des programmes américains dont les objectifs sont semblables, les PRE apparaissent bien peu intensifs à la fois en temps et en moyens, ce qui s'observe déjà par

l'importance de l'investissement financier. Alors que le programme américain *Head Start* consacre environ 8 000 dollars par enfant et par an, le coût d'un PRE est d'environ 850 euros par enfant et par an. Les PRE coordonnent des dispositifs existants sur le territoire mais ne proposent pas réellement d'intervention nouvelle. Ils répondent de façon générique à des besoins spécifiques. De plus, certaines actions du PRE visant à améliorer les résultats scolaires ont été évaluées par ailleurs et ne présentaient pas d'effet positifs significatifs²⁰. Enfin, les effets négatifs observés peuvent refléter l'effet d'étiquetage : le jugement des enseignants n'est pas le reflet de la réalité mais une construction fondée sur des informations diverses (Bressoux & Pansu, 2003) qui agit ensuite en « prophétie auto-réalisatrice ». Les difficultés de coordination des acteurs peuvent aussi expliquer ces impacts négatifs. C'est notamment ce qu'avaient mis en avant Blatchford, Russell, & Webster (2011) pour expliquer l'inefficacité des maîtres auxiliaires en Angleterre.

On perçoit ici tout l'intérêt de l'évaluation d'impact dans l'investissement social. Chaque année, les PRE mobilisent 100 millions d'euros et concernent 100 000 enfants dont en moyenne, la situation aurait été légèrement meilleure s'ils n'avaient bénéficié du programme.

Mais si ces résultats semblent contre-intuitifs par rapport aux objectifs initiaux, de nombreux programmes ont des effets théoriques ambigus et ont besoin d'évaluations rigoureuses pour déterminer si les effets sont positifs ou négatifs. Par exemple, les transferts monétaires pour les élèves d'origine défavorisée sont utilisés dans de nombreux pays dans le but de lever la contrainte financière qui pèse sur les jeunes et les inciter à s'éduquer davantage. L'efficacité de ces dispositifs n'est pas évidente à première vue car s'ils encouragent les jeunes à s'inscrire, leur capacité à permettre aux élèves de réussir dépend du comportement de ces derniers une fois inscrits. Dans leur analyse de l'effet des bourses sur critères sociaux dans le supérieur en France, Fack et Grenet (2015) exploitent les discontinuités d'accès aux bourses en fonction du revenu des parents et montrent qu'une allocation modeste (1500 euros par an) augmente le taux d'inscription dans le supérieur, la poursuite d'études et la probabilité d'obtenir un diplôme.

À partir de ces quelques travaux, on peut déjà dire que l'investissement social dans l'éducation ne peut pas s'appréhender par une augmentation de ressources, une intervention globale comme des tailles de classes réduites. En l'état, les évaluations d'impact de politiques d'investissement social dans l'éducation offrent des résultats contrastés suivant les dispositifs et aux coûts d'intervention très variables. Le fait que des programmes de soutien aux élèves (que ce soit des programmes de tutorat ou les PRE) produisent des effets négatifs appelle davantage d'évaluation rigoureuse des interventions et des répliques pour renforcer le niveau de preuve.

²⁰ En particulier, le programme « coup de pouce clé » a été évalué à trois reprises par des équipes différentes qui ont montré qu'il n'apportait aucun bénéfice par rapport à une prise en charge standard de l'Éducation nationale (Piquée & Suchaut (2002), Goux, Gurgand, & Maurin (2013), Florin, Guimard, & Nocus, (2014)).

Des revues systématiques de littérature existent pour certaines interventions spécifiques. Par exemple, Wilson, Tanner-Smith, Lipsey, Steinka-Fry, & Morrison (2011) proposent une recension systématique des évaluations de programmes de prévention du décrochage publiées entre 1985 et 2010. Ils retiennent 167 études à partir desquels ils concluent que la plupart des programmes parviennent à réduire le décrochage. Ces résultats sont vrais pour les dispositifs scolaires ou des interventions auprès des communautés.

c) Politiques d'activation sur le marché du travail

Les politiques d'activation sur le marché du travail sont au cœur des politiques européenne d'emploi depuis 30 ans et peuvent être interrogées au regard de l'investissement social. Elles prennent des formes très variées tel que des aides à la recherche d'emploi, des contrôles des demandeurs d'emploi, des formations, des emplois aidés,...

Les politiques qui visent à favoriser le retour à l'emploi permettent de mieux utiliser le capital humain des individus et entrent *a priori* comme un moyen de répondre aux objectifs de l'investissement social. Cependant, dans de nombreux pays (Angleterre, Etats-Unis, Pays d'Europe de l'Est), les politiques d'activation ont plutôt été un prolongement de politiques néolibérales qu'un moyen d'améliorer les compétences des individus et de produire davantage de « bons » emplois (Bonoli G. , 2009) (Bonoli G. , 2012). L'investissement social partage avec l'activation l'idée que l'emploi constitue la première protection contre la pauvreté mais elle s'en distingue en insistant sur l'idée que la participation sociale ne peut aller sans protection des plus vulnérables (Duvoux, 2016). Les politiques d'activation doivent donc être interrogées au regard de leur impact sur le retour à l'emploi, mais aussi sur la sortie de la pauvreté, le revenu et les conditions de vie des populations affectées.

En Grande Bretagne un programme d'ampleur visant au retour à l'emploi de pensionnaires d'invalidités a été mis en œuvre au début des années 2000 : *Pathway to work*. Cette politique est particulièrement intéressante car elle s'est dotée au préalable d'une stratégie d'évaluation remarquable.

Plusieurs séries de résultats ont ainsi été obtenues (voir annexe) grâce à une prise d'informations très en amont du déploiement auprès du public cible, d'une mise en œuvre progressive et d'analyses diverses produites par trois centres d'évaluation différents. Becker, Hayllar, & Wood (2010) analysent les enquêtes menées auprès des bénéficiaires et produisent une analyse descriptive très riche identifiant de nombreuses limites à la reprise d'emploi corréliées aux caractéristiques des bénéficiaires. Knight, et al., (2013) exploitent l'expansion progressive du programme pour estimer par une double différence l'effet de *Pathway to Work*. Ils indiquent que le programme diminue de 2 points de pourcentage la part de bénéficiaires de pensions d'invalidité et des éléments indiquent qu'il affecte positivement l'emploi. Cependant, des problèmes de fiabilité des données et d'imprécision statistique empêchent les auteurs d'être parfaitement affirmatifs de ce point de vue.

Adam, Bozio, Emmerson, Greenberg, & Knight (2008) ont produit une analyse coûts-bénéfices à partir des évaluations des effets du programme par d'autres auteurs et de microsimulations visant à identifier les transferts de contribution sociales suite à la mise en œuvre du programme. Adam, Bozio, & Emmerson (2009) ont montré qu'une des analyses produites ne permet pas d'identifier réellement un effet causal car trop de paramètres inobservables viennent brouiller les résultats.

D'un point de vue pratique, *Pathway to Work* est exemplaire dans sa mise en œuvre et son évaluation, multiple, pluridisciplinaire, pluri-méthodologique. Le programme a été déployé progressivement permettant ainsi de comparer les situations dans le temps et dans les zones affectées et non-affectées.

En France, la vaste majorité des réformes n'ont pas pris ce chemin, limitant les possibilités d'évaluations d'impact d'aussi grande ampleur. Une exception encourageante est à trouver dans le programme *working first* français de la « Garantie Jeunes » destiné aux jeunes ni en emploi, ni en formation et en situation de grande précarité sociale. Ce dispositif porté par les missions locales offre à ses bénéficiaires un accompagnement intense pour l'insertion professionnelle et une garantie de revenu. Ce programme a été déployé progressivement (mais généralisé avant la fin de l'expérimentation) permettant de mettre en œuvre une évaluation quasi expérimentale convaincante, et complétée d'une évaluation qualitative riche. Les premiers résultats sont positifs et importants même si les dernières informations obtenues viennent contrebalancer un peu l'évaluation (les résultats sur l'emploi dans la deuxième vague de territoires, initialement moins défavorisés, sont pour l'instant nuls) (DARES, 2016).

De nombreuses évaluations des politiques d'activation ont été mises en place dans plusieurs pays, à partir d'expérimentations ou de quasi-expérimentations testant différents types d'interventions. Elles ont été recensées dans trois revues systématiques récentes qui différencient par type d'intervention.

Card, Kluve, & Weber (2010) s'intéressent au signe et à la significativité d'évaluations d'impacts de politiques d'activation sur des variables très différentes (allant de la durée d'inscription au chômage, le statut d'emploi aux revenus du travail). À partir de 199 impacts tirés de 97 études parues entre 1995 et 2007, les auteurs montrent que les politiques d'activation ont globalement des effets positifs mais que les résultats dépendent de la variable de mesure et du type de programme mené. Les politiques de création d'emplois publics subventionnés sont relativement inefficaces contrairement aux programmes d'accompagnement à la recherche d'emploi qui produisent plus souvent des effets²¹ positifs. Les formations ont en général un impact négatif ou nul à court terme mais un effet positif à moyen terme. Les résultats obtenus et leurs signes ne sont par ailleurs pas liés au fait que le programme soit expérimental ou basé sur des expériences naturelles et des méthodes quasi expérimentales.

²¹ La méta-analyse groupe les effets sur le statut d'emploi, le revenu, etc.

Filges, Smedslund, Knudsen, & Jørgensen (2015) gardent 39 études parmi 79 répondant à leurs critères de sélection²² et indiquent également des effets globalement significatifs. Les auteurs sont plus réservés sur l'ampleur des effets. Leurs estimations montrent qu'il y a 52 % de chances qu'un chômeur bénéficiaire d'un programme d'activation trouve un emploi avant un autre chômeur non bénéficiaire, soit à peine plus d'une chance sur deux. Pour 15 personnes bénéficiaires d'un dispositif d'activation, une aura un emploi un an après l'entrée dans le programme. Les résultats méta-analytiques sur la qualité des emplois sont mitigés.

Contrairement à Card, Kluve, & Weber (2010), les auteurs ne parviennent pas à montrer qu'un type de programme est plus efficace que les autres. Ces différences s'expliquent principalement par les critères d'inclusion conduisant à ne retenir que 39 études au final, et les méthodes d'agrégation des effets mesurés. L'évaluation des programmes d'activation nécessite d'employer des méthodes économétriques souvent assez sophistiquées lorsque le programme n'est pas expérimental. Heckman, LaLonde, & Smith (1999) discutent les hypothèses, les méthodes et les conséquences des différentes stratégies empiriques employées dans la littérature.

Card, Kluve, & Weber (2015) amendent légèrement leurs précédentes conclusions. À partir de 200 impacts, ils montrent que les effets moyens sont proches de 0 à court terme mais deviennent plus élevés 2 ou 3 années après avoir participé au programme. Les effets les plus importants sont obtenus à partir des programmes de formation et d'emploi dans le secteur privé. Il existe des différences importantes entre les profils de participants. Les femmes et les chômeurs de longue durée bénéficient davantage des programmes d'activation. Plus surprenant, les auteurs concluent que les politiques d'activation ont plus de chances d'être efficaces pendant les périodes de récession²³.

En France, Cavaco, Fougère, & Pouget (2009) estiment que la comparaison de personnes semblables qui ont été licenciées mais dont certaines ont pu bénéficier d'une « convention de conversion » ont une probabilité 6 points plus élevée d'être en emploi deux à trois ans plus tard. Cette analyse ne permet toutefois pas d'éliminer parfaitement les différences inobservables entre le groupe bénéficiaire de la formation et les autres. Crépon, Ferracci, & Fougère (2012) ont une stratégie plus robuste et montrent que si la formation des chômeurs augmente la durée de la période de chômage, elle augmente également la durée de l'emploi futur, ce qu'ils interprètent comme le signal d'une meilleure concordance employeurs-employés permise par la formation.

Des expérimentations d'accompagnement des chômeurs ont été menées et leurs résultats sont synthétisés par Parent, Sautory, & Desplatz (2013). Elles ont mis en exergue des effets positifs sur la reprise d'emploi qui tendent toutefois à disparaître à moyen terme. Behaghel, Crépon, & Gurgand (2014) ont par ailleurs montré que le suivi

²² Principalement avoir un groupe de contrôle bien défini.

²³ Ce résultat est étonnant car les périodes de récessions sont typiquement caractérisées par une pénurie d'offres d'emplois ; les politiques d'activation ne peuvent rien au manque de demande de travail à l'exception des créations directs d'emplois (fonction publique ou emplois aidés).

des chômeurs est aussi efficace qu'il soit proposé par une société privée ou par Pôle emploi. Néanmoins, Crepon, Gurgand, Kamionka, & Lequien (2012) indiquent à partir des mêmes données que le coût de l'aide à la reprise de l'emploi est très élevé pour les bénéfices obtenus. En outre, les programmes de soutien non universalistes de ce type peuvent avoir des effets indésirables sur les personnes non éligibles. Crépon, Duflo, Gurgand, Rathelot, & Zamora (2013) exploitent l'organisation d'une expérimentation de soutien à la recherche d'emploi auprès des jeunes dont le design modifie au hasard le taux de personnes soutenues sur les territoires du groupe expérimental. Ils montrent alors que le programme a un effet positif sur l'accès à l'emploi mais que dans les zones traitées intensément, pour les hommes, cet effet se fait au détriment des individus du groupe de contrôle. Les bénéficiaires de l'accompagnement à la recherche d'emploi ont tendance à évincer les chômeurs éligibles mais non bénéficiaires, faisant ainsi subir une externalité négative aux autres chômeurs. Le programme évalué semble alors déplacer les chômeurs dans une forme de file d'attente et n'aurait pas d'effet à long terme²⁴.

Ces nombreux travaux offrent déjà une perspective intéressante pour l'investissement social et son évaluation. Ils montrent bien l'intérêt des évaluations précises et fines de l'efficacité des différents dispositifs mis en œuvre, mais des recensions exhaustives et analytiques sont nécessaires. Certains projets se sont avérés très efficaces, d'autres moins, voire affectent négativement les bénéficiaires ou d'autres par des effets externes. Certains sont extrêmement coûteux, d'autres ont un coût d'intervention quasi nul. Aussi l'impact des politiques d'investissement social doit-il être mis en perspective avec les coûts.

2. Des effets de l'investissement social à son rendement

Le nombre d'évaluations rigoureuses de politiques assimilées à de l'investissement social (dans cette note) s'est beaucoup accru ces dix dernières années ce qui offre des perspectives intéressantes pour de nouvelles interventions. Néanmoins, il est toujours très difficile pour les décideurs de comparer les résultats de différents programmes et leur évaluation lorsqu'ils ont été menés dans des contextes ou pays différents, à des moments distincts, en employant des dispositifs différents pour atteindre un même objectif. La multiplication d'évaluations, souvent complexes, peut donc être un frein à leur utilisation par les décideurs. C'est pourquoi il est intéressant de proposer des

²⁴ Cet argument est une critique forte des programmes de soutien aux demandeurs d'emplois mais un amalgame est souvent fait entre le programme et l'expérimentation. En effet, beaucoup voient dans ce résultat la manifestation du manque de validité externe de la démarche expérimentale : la généralisation de l'expérimentation n'aura aucun effet car la source d'effet provient du non-universalisme de l'expérimentation. En réalité, ce n'est pas l'expérimentation mais l'intervention qui ne peut être généralisée pour ces raisons même si la critique reste valide.

indicateurs plus synthétiques permettant de comparer l'efficacité de différentes interventions en les rapportant à leurs coûts.

2.1. Les approches hypothétiques

Une première approche pour quantifier le rendement de l'investissement social consiste à comparer des situations contingentes dans lesquelles on définit, par hypothèses, les effets de l'investissement social²⁵. Ce type d'analyse part généralement de statistiques descriptives pour caractériser les situations avec ou sans investissement social – revenu moyen des femmes en emploi, performances des enfants qui sont allés en maternelle vs celles de ceux qui n'y sont pas allés, etc. Ces situations sont ensuite comparées en faisant l'hypothèse – explicite ou non – que ces différences sont dues à l'investissement social. Dit autrement, le contrefactuel est défini par hypothèses, il n'est pas identifié empiriquement.

Esping-Andersen & Pallier (2008) calculent de cette façon le rendement des modes de garde en considérant l'amélioration hypothétique du taux d'emploi des femmes que les auteurs attribuent au mode de garde. Les auteurs comparent alors le coût par enfant de 2 ans de crèche et 3 ans de maternelle aux gains salariaux de la mère qui peut alors travailler. Plus spécifiquement, les auteurs se placent dans le cadre danois et considèrent une femme âgée entre 30 et 35 ans, mère de deux enfants qui ne s'arrête de travailler qu'une année pour sa maternité et part en retraite à 60 ans. La situation est comparée avec celle où cette même mère, *parce qu'elle n'aurait pas accès au mode de garde proposé grâce à la stratégie d'investissement social*, n'aurait pas travaillé pendant 5 ans (et subit donc un coût d'opportunité de 5 ans de salaire à taux plein), et aurait perdu en employabilité de sorte qu'en cumulé sur le cycle de vie, son salaire aurait été de 45 % inférieur. La dépense par enfant pour le mode de garde est de 72 850 euros, 5 ans de salaires actualisés rapportent 114 300 euros, la non-perte de salaire sur le cycle de vie actualisé rapporte 200 100 euros. Le rendement social sur le cycle de vie sous ces hypothèses apparaît très élevé : chaque euro dépensé dans le mode de garde rapporte 4,3 euros de salaire à la mère et même un effet net sur les finances publiques de 1,3 euro rapporté par euro dépensé. Bayet (1996) avait proposé des calculs semblables avec des hypothèses légèrement plus sophistiquées (voir Chauffaut & Boyer (2010) pour une discussion sur ces travaux). Maurin et Roy (2008) utilisent les données de l'enquête *Familles et modes de garde* menée à Grenoble et les statistiques descriptives issues de cette enquête pour calculer le rendement d'une place en crèche en ne considérant que les trajectoires de salaires des mères. Cependant, si l'accès au mode de garde ne diffère pas en fonction des caractéristiques observables, on peut douter de la comparabilité des groupes. On peut tout à fait imaginer des caractéristiques inobservables (réseau, localisation et comportement stratégique, négociation, etc.) qui permettent à certains couples d'avoir une place et sont liées aux perspectives d'emploi des mères. L'analyse ne tient pas compte par ailleurs des choix et des caractéristiques du conjoint.

²⁵ Nous ne parlons pas ici d'hypothèses *ex-ante* mais d'estimation du rendement de politiques existantes.

Ce type de calculs de rendement a une visée pédagogique et facilite la compréhension de ce qui peut constituer les sources d'économies ou de richesse future créée. Mais ces travaux s'appuient sur des hypothèses à la discrétion de l'analyste qui peuvent être très éloignées de ce que serait la réalité sans l'investissement social. Le cas du calcul d'Esping-Andersen (2009) constitue un parfait exemple. Dans un monde où les modes de garde n'existeraient pas, l'hypothèse que les femmes ne travailleraient pas pendant 5 ans n'est pas une bonne approximation de la situation contrefactuelle. La réponse des femmes sur le marché du travail à la disponibilité d'un mode de garde est mise en exergue dans de nombreuses études, mais elle n'est pas totale sur toute la période. Quand l'enfant s'autonomise, le retrait porte davantage sur la marge intensive²⁶ (Lefebvre, Merrigan, & Verstraete, 2009) (Bauernschuster & Schlotter, 2015). De plus ce calcul néglige le recours aux modes de gardes informels (familles, amis, voisins, etc.) et surestime les gains.

Nous ne développons pas plus ce point.

2.2. L'apport de l'évaluation socioéconomique appliquée aux politiques sociales

L'analyse économique a développé tout un champ consacré à l'évaluation des projets et politiques publiques afin d'en estimer le rendement. De façon très simplifiée, il existe deux formes de travaux : l'une, plutôt normative et très formalisée, s'appuie sur des modèles économiques calibrés à partir de données empiriques ; l'autre, plus comptable, identifie et projette la valeur des différents coûts et bénéfices. Historiquement, ces analyses ont été largement employées pour l'évaluation de grands investissements publics dans les secteurs des transports, de l'énergie, de l'environnement et du développement durable. Plus récemment, les analyses socioéconomiques se sont répandues assez massivement dans le secteur de la santé et des applications se multiplient pour les politiques sociales.

Cependant, ce développement ne se fait pas sans difficultés. Pour Vining et Weimer (2010), l'application de calculs coûts-bénéfices aux politiques sociales doit répondre à quatre enjeux majeurs :

- il y a besoin d'une approche intégrée de l'évaluation de l'effet des politiques sociales et des calculs coûts-bénéfices ;
- il est nécessaire de reconnaître et d'intégrer explicitement la grande incertitude dans les prévisions et les valorisations qu'implique ce type d'analyse ;
- les comportements qui ne satisfont pas aux hypothèses des modèles économiques usuels doivent être pris en compte et traités en tant que tels ;
- l'application à des politiques qui ont un effet redistributif important doit être adapté à cette spécificité ; le calcul socioéconomique ne s'intéressant en principe qu'à

²⁶ Le nombre d'heures travaillées hebdomadaires.

l'efficience du projet, ce qui est indépendant des questions de distribution des richesses.

Les deux premiers points sont particulièrement importants pour les analyses coûts-bénéfices (et coût-efficacité que nous discutons en détail ci-après), alors que les deux suivants concernent davantage les analyses plus normatives.

En première lecture, le caractère exhaustif et intégré de l'évaluation peut sembler trivial et absolument pas spécifique au champ social. Les analyses coûts-bénéfices posent pourtant davantage de problème dans ce champ car beaucoup de leurs effets sont difficiles à identifier (voir première partie sur l'évaluation d'impact de l'investissement social), interagissent, s'étalent sur une période longue sans que l'on puisse les identifier clairement. Il y a souvent de nombreux effets externes non marchands, qui affectent différents domaines. Par exemple, les investissements dans l'éducation et la santé ont des effets croisés qui, *in fine*, se répercutent sur le capital humain et la participation au marché du travail. Les effets des politiques sociales peuvent également persister pendant une période longue mais sont difficiles à prédire.

Cette difficulté est renforcée par l'incertitude et l'imprécision des estimations elles-mêmes, et multipliée par la nécessité d'apposer une valeur monétaire à ces effets.

L'application de calculs socioéconomiques à des investissements sociaux consiste principalement à calculer les bénéfices nets de la politique à partir d'études d'impacts nombreux et estimés de façon imprécise, parfois à la limite de la significativité statistique, qu'il faut valoriser en unité monétaire à partir de prix implicites incertains. Au fond, ce type de calcul fait la somme de chiffrages largement aléatoires, dimension qu'il est nécessaire d'intégrer dans les calculs, sans quoi les valeurs obtenues peuvent en réalité cacher une grande volatilité et une rentabilité hasardeuse.

Or, prédire des bénéfices nets en utilisant la moyenne de variables aléatoires ne produit pas nécessairement une estimation efficace²⁷ du bénéfice net moyen (Vining & Weimer, 2010). Une façon de tenir compte de l'incertitude sur les effets et leur valorisation monétaire est d'avoir recours à des simulations de Monte Carlo.

Ces simulations consistent à considérer les estimations des effets et leur valorisation monétaire comme les réalisations de variables aléatoires dont on peut spécifier les deux premiers moments (espérance et variance) et éventuellement la forme de la distribution théorique sous-jacente, et à reproduire des tirages de ces distributions. On peut donc ensuite répliquer les calculs socioéconomiques de coûts-bénéfices à partir de centaines de répliqués de tirages aléatoires des distributions des effets et des valeurs associées et ainsi, obtenir une distribution des valeurs nettes, ratio coût/bénéfice, etc. (voir *infra*).

²⁷ Au sens statistique du terme.

2.3. Des analyses coûts-efficacité aux analyses coûts-bénéfices

Nous présentons ici plus en détail cette forme de calcul socioéconomique qui lie les résultats d'études d'impact à leurs coûts.

2.3.1. Définir le ratio coût/efficacité

Le ratio coût/efficacité appliqué dans le champ de l'investissement social est une façon relativement simple et transparente de calculer des formes de rendement non pas en terme monétaire, mais sur une mesure de l'efficacité du programme (par exemple, le nombre d'euros nécessaire pour augmenter le score moyen des élèves d'une unité). Il offre une vision du retour sur investissement pour une situation donnée et résume des interventions complexes en un ratio simple, une mesure commune qui permet de comparer différents programmes.

L'analyse coût-efficacité calcule le ratio de la « quantité d'effet » qu'un programme produit pour un montant donné ou, réciproquement, le coût nécessaire pour atteindre un niveau d'impact donné (Dhaliwal, Duflo, Glennerster, & Tulloch, 2012).

Formellement, le ratio coût/efficacité est : $CE = \frac{\text{Coût total de la politique}}{\text{Impact total de la politique}}$

Ce ratio s'interprète comme le coût pour améliorer le résultat d'une unité (par exemple, le coût par crime évité, le coût par point de score gagné à un test standardisé), mais son inverse est parfois également proposé. L'inverse de ce ratio indique l'impact permis par unité monétaire dépensée.

Suivant les résultats, il est plus compréhensible d'utiliser un ratio que l'autre. Dire que 1 000 dollars dépensés permettent d'éviter 0,32 crime est peut-être moins informatif que de dire que le coût évité d'un crime grâce au programme est de 3 100 dollars (Algan, Beasley, Vitaro, & Tremblay, 2016).

Les analyses coût-efficacité (ACE) peuvent être utilisées *ex-ante* si une estimation de l'efficacité de la politique existe – à partir d'une évaluation pilote ou d'un programme semblable mis en place dans un contexte similaire – ou *a posteriori* pour évaluer le rendement de l'investissement social mené.

Il existe une littérature abondante sur la manière de mener des analyses coût-efficacité ou coût-bénéfice dont la majorité a été écrite pour évaluer les politiques américaines (U.S. Department of Health and Human Services, 1996) (Levin & McEwan, 2001). Mais il n'existe pas de méthode unique pour réaliser ces analyses. Dans chaque situation, il est nécessaire de faire des choix et des hypothèses qui peuvent avoir une influence forte sur les résultats. Ces choix dépendent notamment des objectifs des décideurs ou des acteurs qui peuvent chercher à valoriser le ratio coût/efficacité pour eux ou pour la société dans son ensemble. Par exemple, l'inclusion des coûts pour les utilisateurs dans l'analyse est souvent sujette à débat. Une difficulté est alors de trouver l'équilibre entre le besoin de présenter des comparaisons suffisamment intuitives et compréhensibles

pour les décideurs, et plus généralement les parties prenantes, et une information suffisante pour leur permettre de saisir les nuances et finesses du programme, et notamment la sensibilité aux hypothèses et à l'influence de certains facteurs importants.

La clé de la crédibilité de l'analyse coût-efficacité réside dans l'explicitation et l'argumentation systématique des choix et des hypothèses formulées et de calculs de sensibilité des résultats aux hypothèses alternatives.

1) Quantifier les impacts

Les sources d'évaluation d'impacts sont très nombreuses et vont de comparaisons simples avant/après aux résultats d'expérimentations aléatoires contrôlées dotées d'une grande validité interne. Pour les mêmes raisons que nous avons recommandé d'axer davantage l'évaluation de l'investissement social sur les travaux qui attachent une attention toute particulière à la possibilité d'interpréter leurs résultats de façon causal, nous pensons qu'il est crucial de n'utiliser que ces résultats dans les ACE.

Le Washington State Institute for Public Policy (2016) considère au minimum des analyses dans lesquelles les individus affectés et non affectés sont appariés par des méthodes statistiques sur au moins quelques variables de contrôle. Cependant, il recense de façon systématiques les travaux et utilise l'effet agrégé – un coefficient méta-analytique – dans l'analyse.

La précision des estimations doit également être prise en compte dans l'analyse coût-efficacité. En général, la valeur de l'estimation paramétrique est utilisée dans les analyses, ou du moins lorsque celle-ci est significativement différente de 0. Cependant, les ACE visent généralement à discriminer entre différentes interventions. Or, il est possible que les valeurs des impacts apparaissent différentes mais que leur précision ne permette pas de les distinguer en raison de l'imprécision des estimations. Ces comparaisons doivent tenir compte de la précision en présentant par exemple des intervalles de confiance.

2) Le cas des impacts multiples

Les analyses coûts-efficacité sont par définition focalisées sur un résultat unique. Pourtant, la plupart des politiques d'investissement social ont, directement ou indirectement, des impacts variés sur plusieurs dimensions. Ainsi, le choix d'un seul indicateur peut ne pas refléter l'intégralité du programme. Une politique d'accès aux modes de garde peut avoir un effet sur l'emploi des femmes mais également un effet sur le développement de l'enfant. Deux effets qui sont obtenus à partir des mêmes euros dépensés.

Les ACE ne sont pas particulièrement adaptées à ce type de situations. Les analyses coûts-bénéfices décrites plus bas peuvent mieux répondre à ces problèmes même si elles reposent sur des hypothèses plus fortes. Cependant, dans certains cas, il peut être possible de ventiler les coûts par effets attendus.

Dans l'exemple précédemment évoqué, pour un programme qui consisterait d'un côté à accueillir les enfants en bas âge et de l'autre à faire intervenir des personnels formés pour organiser des activités d'éveil, il peut être raisonnable de faire correspondre le coût d'accueil à l'effet sur l'emploi des mères et celui spécifique aux intervenants aux effets sur le développement des enfants.

Une alternative possible consiste à prendre la politique d'investissement social comme un ensemble et de l'évaluer comme tel. Dans ce cas, on peut envisager d'allouer le coût total entre les différents effets identifiés. Si le ratio coût/efficacité pour tous les résultats est supérieur à la meilleure alternative pour obtenir ce résultat, alors le programme est manifestement coût-efficace. Dans leurs travaux, Miguel & Kremer (2004) analysent l'effet d'un programme expérimental de déparasitage²⁸ des enfants qui s'est révélé très efficace sur la santé des enfants et sur leur assiduité scolaire. En divisant le coût du déparasitage en deux pour ces deux résultats, Dhaliwal, Duflo, Glennerster, & Tulloch (2012) indiquent que les ratios coûts/efficacité du programme sont très élevés et que le déparasitage est une intervention très coût-efficace dans son ensemble.

3) Quantifier les coûts

L'évaluation des coûts de la politique d'investissement social peut être extrêmement compliquée à mettre en œuvre au-delà d'une estimation très agrégée. La plupart des travaux publiés évaluant des politiques – que nous identifions à de l'investissement social – ne fournissent pas assez d'information sur les coûts spécifiques des interventions pour mettre en œuvre une analyse coût-efficacité comparée. Il est alors souvent nécessaire de les reconstituer à partir de sources alternatives, ce qui risque d'introduire des erreurs de mesure. On peut avoir recours à la méthode dite des « ingrédients » pour reconstituer les coûts (Levin & McEwan, 2001). Ce faisant, il faut définir clairement ce que représentent le programme et le contexte dans lequel il est amené à être répliqué afin de ne pas mélanger les coûts qui seront supportés de toute façon et ceux qui incombent réellement à la politique mise en œuvre. Parfois, une politique est totalement neuve et tous les coûts sont à inclure, parfois ce ne sont que des ajustements à la marge qu'il faut intégrer. Le problème, c'est qu'il n'y a pas de façon unique de procéder à ces choix.

Calculer le coût d'une politique d'investissement social nécessite donc de faire des choix et des jugements sur ce qui constitue un coût, ce qui doit être inclus ou non et à qui ce coût incombe. Ces choix peuvent ne mobiliser que les ressources additionnelles²⁹ ou être plus exhaustifs. Les choix dépendent de la question de recherche.

²⁸ La contamination de l'eau dans les pays en développement par des parasites est la source d'une morbidité importante particulièrement virulente chez les enfants. L'expérimentation a ainsi consisté à fournir des vermifuges et autres antiparasites aux enfants.

²⁹ J-PAL considère généralement les coûts marginaux, considérant qu'en l'absence d'intervention, la plupart des coûts fixes seraient de toute façon présents. Par exemple, dans le cas d'un programme de formation des enseignants, J-PAL ne considère que le coût de la formation car en l'absence de programme,

Pour nous, le choix généralement pertinent de valorisation devrait être la différence entre les coûts totaux et les coûts qui prévalent dans la situation contrefactuelle, ce qui revient à adopter le coût marginal préconisé par J-PAL. Ce choix est d'autant plus légitime que l'évaluation est basée sur le contrefactuel. Une autre règle d'inclusion peut être celle du passage à l'échelle ou de la réplique. Dans ce cas, certains coûts fixes peuvent être pris en charge (Si on s'intéresse au ratio coût/efficacité d'une diminution de la taille des classes au primaire dans l'éducation prioritaire, il est utile d'intégrer les coûts d'utilisation du mobilier, de l'immobilier et les coûts administratifs car il est probable qu'un passage à l'échelle nécessite de nouveaux locaux, etc.).

La situation de référence est donc un élément clé qui doit être explicitement définie et commune pour les différentes alternatives de politique dont on compare le ratio coût/efficacité.

La prise en compte des coûts pour les utilisateurs (ou bénéficiaires) peut sembler hors de propos du point de vue du décideur politique. Par exemple le temps passé en formation pour un chômeur représente un coût d'opportunité dont la valorisation est complexe mais qui est directement liée à l'investissement dans la formation. Il est utile de l'inclure dans la valorisation des coûts si le calcul vise à mesurer le rendement social. Si au contraire la question est de savoir si le financeur de projet percevra un retour sur investissement, ce coût peut être écarté.

D'une manière plus générale, l'estimation du coût d'une politique d'investissement social est fortement liée à l'évaluation contrefactuelle puisqu'il s'agit de mesurer ce qui est dépensé en plus, impliquant implicitement une situation de référence. Nous illustrons ce point par l'évaluation du coût du redoublement. Sans aller jusqu'à une analyse coût-efficacité, de nombreuses études ont estimé ce coût en faisant le produit de la dépense moyenne par élève et le nombre d'élèves ayant redoublé. Implicitement (ou non), ce calcul prend pour référence une situation où il n'y aurait pas de redoublement et suppose que le redoublement allonge la durée de scolarité de précisément une année. Or, plusieurs travaux ont montré que le redoublement augmente la probabilité de décrocher et affecte les trajectoires scolaires dans bien des manières (Heim, Seinmetz, & Tricot, 2015), mettant en doute cette hypothèse et la valorisation du coût qui en découle. Pour en tenir compte, Benhenda et Grenet (2015) ont réalisé une estimation de l'effet du redoublement sur les trajectoires scolaires avant d'en estimer le coût. Constatant que les élèves nés en fin d'année ont une probabilité plus élevée de redoubler que ceux nés en début d'année, ils utilisent le mois de naissance comme instrument expliquant la probabilité de redoubler. Leur analyse est donc explicitement fondée sur la valorisation du coût du redoublement par rapport à une situation où le redoublement n'existerait pas et ses effets sur les trajectoires scolaires non plus. L'apport de leur calcul tient également à la prise en compte des effets dynamiques pour faire apparaître les économies que générerait la suppression du redoublement.

les coûts pour l'accueil des enfants, l'utilisation des infrastructures et les salaires du personnel éducatif seraient de toute façon dépensés.

4) Utiliser des mesures communes

Un obstacle souvent mal appréhendé de l'ACE, au-delà de la sélection des bonnes mesures de l'impact et l'inclusion des coûts pertinents, est de convertir tous les coûts et les impacts en unités communes, en tenant compte de l'inflation, éventuellement des taux de changes et des années de mise en place.

Il n'est pas utile dans ce travail d'aller au-delà d'alerter sur ces enjeux qui doivent être traités au cas par cas. Le lecteur intéressé peut se référer à la section de Dhaliwal, Duflo, Glennerster, & Tulloch (2012) dédiée pour une discussion plus complète.

2.3.2. De l'efficacité au bénéfice : l'inflation des hypothèses

L'analyse coûts-efficacité montre la quantité d'effet qu'un programme produit pour un coût donné ; l'analyse coûts-bénéfices (ACB) combine l'ensemble des bénéfices d'un programme en une échelle unique, généralement monétaire, et présente le ratio des bénéfices combinés et du coût. L'avantage de l'ACB est de mieux appréhender les programmes qui affectent différentes dimensions. En outre, ramener les coûts et les bénéfices en une même unité monétaire permet de passer d'une logique relative – en comparant les ACE de différentes alternatives – à un jugement « dans l'absolu ». Il est alors possible de juger de la pertinence d'un investissement en fonction de son rendement monétaire.

Les analyses coûts-bénéfices sont également appelées calculs socioéconomiques. Elles sont particulièrement intéressantes dans des situations où un investissement initial doit générer des flux de bénéfices futurs et donc dans le champ de l'investissement social. Le ratio coûts/bénéfices n'est pas le seul indicateur de rendement. Il est souvent intéressant de présenter le taux de rentabilité interne, taux de croissance moyen qui égalise la somme des coûts et bénéfices actualisés, ou la valeur actuelle nette, somme actualisée des coûts et bénéfices.

Comme pour les ACE, les ACB peuvent être mises en place *ex-ante* afin d'anticiper la rentabilité sociale d'un projet³⁰, ou *ex-post* et permettre d'évaluer le rendement social de l'investissement mené précédemment. Un exemple emblématique et remarquable de rigueur et de transparence est à trouver dans Heckman, Moon, Pinto, Savelyev, & Yavitz (2010). Les auteurs évaluent le rendement du programme d'intervention précoce Perry-Preschool program (voir plus haut) en proposant une solution statistique aux problèmes de données manquantes, d'attrition et même d'altération du protocole expérimental³¹. Les coûts du programme sont pris au sens large et intègrent à la fois les dépenses opérationnelles (enseignants et coûts administratifs) et les coûts du capital (pour les salles et les bâtiments). Ce coût représente 17 759 dollars par enfant (valeur 2006).

³⁰ À notre connaissance, il n'y a pas eu d'évaluation prospective d'investissements sociaux basée sur des ACB rigoureuses.

³¹ Dans un autre article (Heckman J. , Moon, Pinto, Savelyev, & Yavitz, 2010), les auteurs ont montré que la randomisation a été compromise et ont proposé une méthode permettant de corriger ce problème. Ils montrent alors que les estimations de l'effet du programme sont robustes malgré tout.

Mais l'analyse intègre également aux coûts les dépenses d'éducation successives en se fondant sur l'idée que le programme améliore l'éducation à la fois en prolongeant la durée des études et en accélérant l'accès à un niveau d'études donné. Les bénéfiques sont d'abord constitués par les profils de revenus tout au long de la vie (ce qui nécessite des hypothèses pour être estimé), les taxes et cotisations collectées, les prestations sociales évitées mais également par les coûts des crimes évités³². Leurs estimations indiquent alors un taux de rendement interne³³ de 7 % à 10 %, statistiquement significatif. Les ratios coûts/bénéfices proposés en prenant une estimation basse du coût des meurtres est de 7.1 au global, soit de 8.6 pour les hommes et 4.5 pour les femmes. Ainsi chaque dollar investi dans les hommes du programme PPP rapporte à la société 8,6 dollars.

Le problème de cette démarche est qu'elle nécessite un nombre important d'hypothèses pour *monétiser* les bénéfiques pour lesquels différentes organisations peuvent avoir une vision très différente. La simple valorisation monétaire de gains de performance scolaire n'est pas une question triviale. Si de nombreuses études produisent une estimation du rendement d'une année d'études supplémentaire, celles qui estiment les gains de revenus associés à un score en mathématiques ou en littérature plus élevé sont non seulement plus rares mais peuvent difficilement s'interpréter de façon causale.

Le *think tank* britannique New Economy s'est attaché à produire une base de données des coûts sociaux disponible gratuitement³⁴ en ligne.

Cette base de données recense, dans un document unique, les chiffrages de différents coûts (ou coûts évités) pour la société d'interventions ou de non-intervention dans différents domaines (éducation, crime, etc.). Ces chiffrages sont obtenus à partir de travaux académiques ou de publications d'institutions diverses (organisations nationales et internationales, *think tanks*, associations, etc.). Les coûts proposés sont datés mais concernent des unités de mesure différentes (certaines sont par individu par an, d'autres par individu sur toute une vie, d'autres par événement évité, etc.). Les sources sont présentées, les coûts décrits dans une cellule de commentaire. La base couvre (aujourd'hui) les champs des crimes, de l'éducation et des compétences, de la santé, de l'économie et de l'emploi, des incendies, du logement et des services sociaux. D'après les ressources disponibles sur le site, cette base a vocation à être mobilisée dans des analyses coûts-bénéfices principalement *ex-ante* pour évaluer le rendement anticipé de programmes visant à réduire ces coûts. L'intérêt de cette base est alors d'avoir une référence unique.

Les coûts évalués ici sont spécifiques au Royaume-Uni et ne peuvent être directement utilisés dans des contextes nationaux différents.

³² Ce qui a été beaucoup discuté dans la littérature critique sur ce sujet. Voir Lochner (2011) pour une synthèse.

³³ Le taux de rendement interne est le taux d'intérêt qui annule la valeur nette présente. Il est à comparer avec le rendement d'autres actifs financiers.

³⁴ <http://neweconomymanchester.com/our-work/research-evaluation-cost-benefit-analysis/cost-benefit-analysis/unit-cost-database>.

Les calculs coûts-efficacité laissent les lecteurs et utilisateurs de l'analyse la liberté de valoriser les bénéfices à leur façon. L'ACE informe sur les gains à attendre pour un coût donné et laisse à l'utilisateur ou au décideur politique la liberté de juger si ce bénéfice est suffisant. Ainsi, dans le champ de l'investissement social, une évaluation coût-bénéfice nécessiterait entre autre d'estimer les gains de productivité permis par une amélioration de la qualité ou de la quantité d'éducation proposée à la population. Il y a peu de valeurs monétaires conventionnelles, admises et utilisées pour valoriser des gains tels que les années de vie gagnées, des gains de scores à des tests éducatifs, etc., ce qui rend les analyses coûts-bénéfices complexes dans le champ de l'investissement social.

Dans leur article de 2010, Vining et Weimer (2010) tirent le bilan de l'ouvrage qu'ils ont dirigé (Weimer & Vining, 2009) et identifient des besoins de recherche pour déterminer la valeur sociale de plusieurs éléments.

D'abord, certains coûts implicites ont besoin d'estimations plus précises et adaptées au contexte. Il s'agit en particulier de la valeur des crimes et du coût d'opportunité des fonds publics. En France, le seul travail (à notre connaissance) visant à évaluer le coût social des différents crimes et délits est le rapport Bichot (2010). Ce rapport est, de l'aveu de l'auteur, une première étape vers une valorisation systématique mais implique d'engager d'autres recherches. De nombreuses valeurs tutélaires ont toutefois pu être proposées dans ce travail. Par exemple, le coût d'une blessure volontaire est évalué à environ 34 750 euros₂₀₁₀ qui est basé sur une estimation du nombre d'agressions réalisé multiplié par le coût d'un accidenté de la route que l'auteur utilise comme *proxi* du coût social d'une blessure physique à partir d'analyses de l'ONISR³⁵.

Aux États-Unis, les travaux de Miller *et al.* (1996) ont été particulièrement influents en valorisant le coût privé d'un crime avec le coût du dommage ou de la perte de biens, des soins médicaux et psychologiques, le coût d'intervention des forces de l'ordre, des services aux victimes et la perte de productivité des victimes. Les coûts intangibles sont basés sur les dispositions marginales à payer révélées dans des travaux d'économie comportementale (e.g. ce que les gens sont prêts à verser pour éviter une douleur physique, réduire leur risque d'agression ou de décès).

Vining et Weimer (2010) interrogent aussi la valorisation de gains de productivité. Cette question est particulièrement importante dans le champ de l'investissement social car elle découle directement de l'investissement en capital humain. De façon similaire, l'évaluation de la valeur d'années d'études supplémentaires a besoin d'être mieux appréhendée, notamment dans ses aspects non monétaires (tels que les effets sur le comportement en tant que parents, sur la santé, sur la citoyenneté et la participation électorale, etc.). La valeur du développement cognitif n'est pas bien établie.

³⁵ Disponible dans les bilans annuels. En 2011, la valeur d'un blessé hospitalisé est estimée à 132 367 euros.

En bref, les principales valeurs des résultats affectés par les politiques d'investissement social ne sont aujourd'hui pas identifiées de façon consensuelle et inclusive.

Il existe cependant des travaux pour lesquels des calculs coûts-bénéfices peuvent être réalisés sans avoir besoin de formuler trop d'hypothèses. Nous en présentons quelques-uns dans la section suivante.

L'intérêt de ces calculs ne réside pas dans le chiffrage en soit, mais dans la comparaison des ratios coûts/efficacité ou coûts/bénéfices pour différentes alternatives de politiques visant à répondre au même objectif et utilisant les mêmes conventions de valorisation.

2.3.3. Quelques analyses coûts-efficacité et coûts-bénéfices associées à des évaluations

Plus tôt, nous avons présenté les travaux de Fack & Grenet (2015) qui estiment de façon très rigoureuse l'impact des bourses sur critères sociaux en France sur l'accès et la persévérance dans l'éducation supérieure. Ils montrent notamment que recevoir une bourse de 1 500 euros par an environ a un effet positif significatif sur la probabilité de poursuivre des études supérieures. Ils font ensuite une analyse coûts-bénéfices simple.

Si on considère 1 000 candidats aux études supérieures demandant des bourses, les auteurs montrent que 783 s'inscrivent effectivement et parmi eux, 49 ne l'auraient pas fait s'ils n'avaient pas été éligibles. C'est la représentation en effectif du résultat de leur estimation économétrique. En prenant 1 000 candidats pour une inscription en master et demandant une bourse, ces chiffres sont respectivement de 770 et 39. Le coût d'amener des étudiants supplémentaires dans le supérieur grâce à la bourse est résumé par les auteurs au coût de la bourse, aux coûts de scolarité et au coût d'opportunité des études représenté par les revenus non perçus pendant la période d'études. Sur les 1 000 étudiants candidats, 783 perçoivent une bourse et s'inscrivent dans le supérieur, ce qui représente une dépense d'environ 1,5 million d'euros. Cette dépense a permis à 49 étudiants d'accéder aux études supérieures. Ainsi, la dépense en bourses nécessaire pour amener un étudiant supplémentaire est d'environ 24 000 euros. Au niveau master, ces calculs conduisent à un coût par étudiant de 31 000 euros. En considérant le taux de survie dans le supérieur, le coût par étudiant est augmenté de 15 400 euros au niveau licence et 22 200 au niveau master. Le ministère de l'Éducation nationale estime le coût d'une année d'études supérieures en moyenne à 10 770 euros par an, soit pour deux ans et actualisé environ 21 200 euros. Les auteurs estiment les coûts d'opportunité liés aux revenus non perçus pour étudier à 40 000 euros par an pour le niveau licence et 50 000 euros par an pour le niveau master.

Au final, le coût social pour amener un étudiant supplémentaire à aller deux ans dans le supérieur grâce à la bourse est estimé à environ 100 000 euros. Au niveau master à 125 000 euros.

Les bénéfices sont limités aux différences de trajectoires de salaires entre différents niveaux de diplôme estimées à partir de l'enquête Emploi (Insee). Les auteurs estiment ainsi qu'avoir au moins deux ans d'université plutôt qu'un niveau bac produit un différentiel de salaire de 13 % en moyenne ; entre un niveau master plutôt que licence, ce différentiel est de 35 %.

En sommant et actualisant³⁶ ces écarts de salaires sur toute la carrière, les auteurs estiment le bénéfice de deux ans supplémentaires à 145 000 euros, et à 425 000 euros au niveau master. Au final, les ratios coûts/bénéfices d'amener un étudiant à deux années d'études supérieures avec les bourses sur critères sociaux sont respectivement de 1.45 et 4.25.

Cette analyse est, de l'aveu des auteurs³⁷, complètement secondaire, ce qui explique pourquoi les calculs sont très arrondis, comme pour refléter l'incertitude de ces montants.

Un travail assez semblable a été réalisé par McNally & Machin (2008) pour l'évaluation du projet *Literacy Hour*. Les auteurs estiment en effet les coûts annuels du dispositif à 2,5 millions de livres sterling, ce qui revient à la somme dérisoire de 25,52 livres par élève par an. Les estimations de l'effet du programme sur les performances en lecture conduisent à une hausse de 8 % d'un écart-type en lecture. Les auteurs utilisent ensuite une étude longitudinale anglaise pour identifier le lien entre salaire à 25 ans et score en lecture. Ils estiment alors que les gains en lecture permis par le programme devraient permettre une hausse du salaire future telle que, cumulée sur 45 ans et actualisée³⁸, un élève qui a bénéficié du programme gagnera entre 1 920 et 5 000 livres sterling de plus que s'il n'en avait pas bénéficié.

Cook, *et al.* (2014) ont évalué une expérimentation proposant du tutorat à des jeunes défavorisés et issus de minorités de classes équivalentes à la 3^e-2^e qui s'est avérée particulièrement efficace. En effet, les performances en mathématiques des jeunes affectés ont progressé de 65 % d'un écart-type du groupe de contrôle. Or, les coûts de l'intervention sont assez faibles : entre 3 000 et 6 000 dollars par enfant et par an, les auteurs comparent les ratios coûts/efficacité de leur intervention à ceux de programmes très connus tel que *Perry Preschool Program*.

Ce programme peut s'apparenter à celui canadien évalué par Algan, Beasley, Vitaro, & Tremblay (2016) évoqué dans les sections précédentes. Ces derniers avaient également proposé des ratios coûts/efficacité pour les principaux résultats de leur analyse. Ils estiment ainsi qu'un crime est évité pour chaque 3 100 dollars investis dans le programme, qu'un redoublement est évité pour 7 600 dollars, etc. Ils comparent ensuite ces valeurs aux coûts pour la société pour évaluer le bénéfice du programme. Tous calculs faits, ils estiment qu'un dollar investi dans cette intervention produit à terme

³⁶ En supposant un taux d'actualisation de 3 % et un taux de croissance annuel du salaire réel de 1 %.

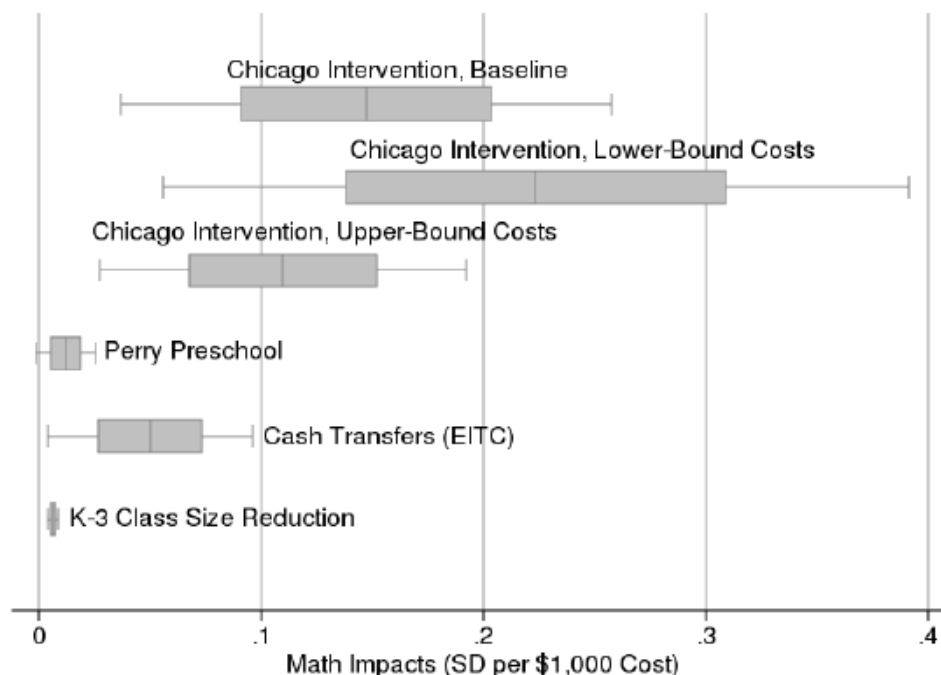
³⁷ Qui parlent de « back of the envelop calculations ».

³⁸ Avec un taux d'escompte de 3 %.

14 dollars de retour pour la société. Néanmoins, les calculs des auteurs ne sont pas très explicites et les coûts ne paraissent pas répartis entre les différents résultats.

Figure 2 – Rapport coûts/efficacité de différents programmes américains sur les performances en mathématiques (d'après Cook et al. 2014)

Figure 6. Impact on Math Test Scores During Adolescence, per \$1,000 program cost (Z-Scores)



Les évaluations coûts-efficacité et coûts-bénéfices sont assez courantes dans les travaux d'évaluation des politiques publiques mais ils apparaissent le plus souvent comme secondaires. Les auteurs les considèrent en général comme des « calculs de coin de table » et passent généralement rapidement dessus³⁹. Il y a cependant des articles focalisés sur l'estimation rigoureuse du rendement de programmes d'investissement social. Nous avons déjà discuté de l'évaluation de *Perry Preschool* par Heckman J., Moon, Pinto, Savelyev, & Yavitz (2010). Dalziel, Halliday, & Segal (2015) proposent une revue systématique des évaluations coûts/bénéfices des programmes d'intervention précoce. Avec leurs critères d'inclusion, ils recensent 13 évaluations économiques concernant six programmes distincts. Ils concluent que sur les six interventions, la moitié présente des bénéfices qui excèdent les coûts ; les autres ne sont pas socialement rentables d'après cette analyse. Les causes de ces écarts sont à trouver d'une part dans les moindres impacts identifiés au départ et au nombre de variables de résultats utilisées et la durée sur laquelle les effets peuvent se manifester.

Les hypothèses formulées pour produire des analyses coûts-bénéfices peuvent donc être très fortes et les résultats extrêmement sensibles à ces choix.

³⁹ Par exemple, Grenet et Fack (2015) parlent de calculs « back of the envelop » et les font figurer en *online appendix* de leur article.

2.3.4. Gérer l'incertitude des impacts et des valeurs par réplication

Si l'évaluation d'impact comporte déjà une grande part d'incertitude dans l'estimation de l'effet d'une politique d'investissement social, les calculs de rendement décuplent considérablement le risque de présenter un résultat aberrant en multipliant entre elles des valeurs très incertaines. Par exemple, le modèle utilisant les valeurs moyennes peut indiquer qu'un programme d'investissement social rapporte à la société 2,50 euros par euro dépensé. La question fondamentale est de savoir à quel point cette estimation est risquée. Si on modifiait au hasard les valeurs des paramètres, est-ce que les bénéfices restent supérieurs aux coûts, ou l'inverse ?

Les simulations de Monte Carlo permettent de répondre à cette question. La méthode consiste à estimer le modèle d'analyse coûts-bénéfices plusieurs milliers de fois en modifiant les paramètres au hasard en échantillonnant des valeurs dans leurs distributions théoriques. Il s'agit alors de conserver ces estimations et à la fin, calculer la valeur moyenne de ces résultats. Il est alors possible d'apprécier le risque de l'estimation en calculant le nombre de fois où les coûts dépassent les bénéfices sur le nombre total de réplications.

Les simulations de Monte Carlo appliquées à l'analyse coûts-bénéfices de politiques sociales présentent aussi un intérêt fort pour communiquer efficacement sur le degré de risque associé aux rendements des projets car on peut présenter la probabilité d'obtenir un bénéfice net positif (comme la fraction des réplications positives) et la valeur moyenne de ce bénéfice.

C'est la méthode adoptée par le Washington Institute for Public Policy, comme l'illustre le tableau ci-dessous.

Figure 3 – Analyses coûts-bénéfices de programmes de développement professionnels et risque d'erreur (Washington State Institute for Public Policy, 2016)

Program name	Date of last literature review	Total benefits	Taxpayer benefits	Non-taxpayer benefits	Costs	Benefits minus costs (net present value)	Benefit to cost ratio	Chance benefits will exceed costs
Case management for unemployment insurance claimants	Nov. 2015	\$3,606	\$1,154	\$2,452	(\$182)	\$3,424	\$19.84	69 %
Training with work experience for adult welfare recipients	Nov. 2015	\$6,928	\$3,781	\$3,147	(\$4,158)	\$2,770	\$1.67	78 %
Job search and placement	Nov. 2015	\$2,345	\$1,491	\$854	(\$517)	\$1,829	\$4.54	67 %
Work experience	Nov. 2015	\$3,834	\$2,570	\$1,264	(\$2,053)	\$1,780	\$1.87	81 %
Training with work experience for adults, not targeting welfare recipients	Nov. 2015	\$5,127	\$2,315	\$2,812	(\$4,138)	\$989	\$1.24	54 %
Case management for former welfare recipients	Nov. 2015	(\$934)	\$391	(\$1,325)	(\$2,936)	(\$3,870)	(\$0.32)	19 %
Case management for welfare recipients or low-income individuals	Nov. 2015	(\$965)	\$246	(\$1,211)	(\$2,919)	(\$3,885)	(\$0.33)	17 %
Training, no work experience	Nov. 2015	\$3,636	\$2,308	\$1,328	(\$8,297)	(\$4,661)	\$0.44	38 %
Training with work experience for youth	Nov. 2015	(\$2,494)	\$675	(\$3,169)	(\$7,362)	(\$9,856)	(\$0.34)	32 %

La figure 4 présente un extrait des évaluations menées pour les politiques d'activation sur le marché du travail. Les revues de littératures ont été menées en novembre 2015. Les bénéfices totaux sont la somme actualisée sur le cycle de vie des bénéfices du

programme monétisé par individu. Par exemple, pour un programme d'aide à la recherche d'emploi, les bénéfices proviennent principalement des salaires retrouvés grâce au programme, à la non-dépense d'allocation chômage ou équivalente et d'autres éléments monétisés. Les coûts sont estimés par individu également. Le rendement est exprimé sous forme de différence ou de ratio. Ainsi le programme de soutien à la recherche d'emploi rapporte 989 dollars de plus qu'il n'en coûte, soit 1,24 dollar de retour par dollar investi. Les simulations pour tenir compte de l'incertitude de ces estimations indiquent toutefois que la probabilité pour que ce résultat soit positif n'est que de 54 %. Par ailleurs, les travaux sur lesquels s'appuie cet exemple particulier semblent un peu datés et peu sont issus de revues à comité de lecture ce qui invite à la prudence

2.4. Retours sociaux sur investissement ou analyses coûts-bénéfices ?

Depuis une vingtaine d'année, la question de l'impact social et environnemental des acteurs économiques est au cœur des préoccupations et a fait naître un outil de mesure appelé « retour social sur investissement ». Celui-ci est « *un cadre d'analyse pour mesurer et rendre compte d'une conception élargie de la valeur incluant les coûts et les bénéfices à la fois sociaux, environnementaux et économiques ; Il vise ainsi à promouvoir les actions participantes à la réduction des inégalités sociales et de la dégradation environnementale ainsi qu'une amélioration du bien-être* » (ESSEC, 2011).

De nombreux travaux ont été mis en œuvre mais le manque de normes de pratiques dans ce champ limite souvent la comparaison des analyses et des résultats et discrédite partiellement la démarche (Olsen, 2003). De nombreuses institutions se sont mobilisées pour produire des guides de bonnes pratiques et d'utilisation, notamment le SROI Network au Royaume-Uni.

La démarche du SROI n'est pas très éloignée de celles des analyses coûts-bénéfices que nous avons décrites précédemment bien que leur cible et leur finalité ne soient pas tout à fait les mêmes. D'après The SROI Network (2012) et l'Essec (2011), les SROI sont basés sur six principes structurants :

- impliquer les parties prenantes,
- comprendre ce qui change,
- évaluer ce qui compte,
- n'inclure que ce qui est « matériel »,
- ne pas surévaluer,
- être transparent,
- vérifier les résultats.

L'implication des partenaires et parties prenantes apparaît comme fondamentale aux SROI alors qu'elle est plutôt proscrite dans la pratique académique des calculs coûts-bénéfices, où l'absence de conflit d'intérêt est centrale pour crédibiliser les analyses. Les

tenants des SROI légitiment l'implication des parties prenantes à ces calculs car ils permettent de limiter « l'omission d'effets importants » (SROI Network, 2015), ce qui peut être interprété de différentes façons... La valorisation des résultats prend le point de vue des parties prenantes. Le choix des mesures n'est donc pas neutre. Nous traduisons ici ce que le SROI Network (2015) identifie comme différences entre les analyses coûts-bénéfices (ACB) et les SROI.

« Les ACB sont généralement utilisées par le secteur public ou quasi public et sont souvent pratiquées avec un niveau élevé de rigueur. Par contraste, les principes des SROI peuvent être utilisés à n'importe quel niveau de rigueur, du moment que c'est « suffisamment bon » pour le type de décision qu'il sert à renseigner. De ce point de vue, les SROI sont semblables à la comptabilité financière, qui vise à fournir une information suffisante et suffisamment précise pour informer d'abord les investisseurs, par opposition aux niveaux de rigueur des sciences sociales. [...] Les évaluations aléatoires contrôlées peuvent être utilisées dans le contexte d'un SROI mais leur utilisation dépend du but et de l'usage de l'analyse. De l'autre côté de l'échelle, l'estimation de la situation contrefactuelle peut être réalisé simplement en demandant aux bénéficiaires ce qu'ils pensent qu'il leur serait arrivé s'ils n'avaient pas participé au programme, sans autre utilisation d'un groupe de contrôle. Ce niveau de rigueur a peu de chance d'influencer un choix politique mais peut être utile pour définir des services ».

Ces quelques éléments montrent bien que les SROI visent à fédérer les investisseurs, financeurs et les autres parties prenantes autour du projet en présentant – un peu par tous les moyens – ce que les actions sont susceptibles de générer. Ils peuvent être un cadre adapté pour comprendre l'action d'une entreprise, d'une ONG, d'une association, sur son environnement (y compris social) mais il ne permet pas d'évaluer de façon rigoureuse l'impact social sauf à s'appuyer sur des estimations d'impacts rigoureuses et une valorisation des coûts adaptés. La lecture du guide du SROI cité précédemment est très claire de ce point de vue.

L'analyse de la pertinence des SROI dans l'évaluation économique de Rauscher, Schober, & Millner (2012) met en cause jusqu'aux choix du lexique dans le SROI – souvent orienté positivement en remplaçant coût par investissement par exemple. Pour les auteurs, la faiblesse des SROI est que « l'analyste doit définir un nombre important de normes pour produire le calcul à sa seule discrétion. Ça s'applique à la fois à la mesure et à l'évaluation de l'impact. »

Les différentes méthodes visant à estimer le rendement reposent sur un nombre d'hypothèses et de normes important qui doivent être explicitement présentées, discutées et testées afin de rendre l'argumentation plus solide. Les SROI apparaissent comme une forme d'analyse coûts-bénéfices moins contrainte par la méthode et plus

conciliante avec les parties prenantes, ce qui n'est pas satisfaisant pour apporter des éléments probants de l'efficacité de l'investissement social.

Ces différents éléments nous offrent une vision plus claire des stratégies d'évaluation possible pour l'investissement social que nous discutons dans la section suivante.

3. Quelles évaluations pour l'investissement social en France ?

L'investissement social a pour objectif de doter les individus en capital humain, de leur permettre de le maintenir et de l'utiliser pleinement. Les stratégies d'investissement social mises en œuvre dans différents pays prennent des formes variées qui ont largement été décrites dans les comparaisons internationales. Cette littérature permet de caractériser les configurations institutionnelles spécifiques et les performances macroéconomiques associées et de souligner les liens et les consistances comparatives. Cette approche est à prolonger pour interroger la pertinence et la cohérence des politiques au regard des objectifs de l'investissement social. Elle ne pourra en revanche être poursuivie pour répondre à la question de l'efficacité des politiques et dispositifs mis en place. L'évaluation de l'effet de l'investissement social doit mobiliser les études d'impact rigoureuses des politiques et interventions visant à augmenter, maintenir ou utiliser le capital humain.

3.1. Pertinence et cohérence des politiques sociales au regard de l'investissement social

Pour faciliter l'argumentation sur les moyens d'identifier l'impact des politiques d'investissement social, ce document a gardé une définition large de l'investissement social compris comme l'ensemble des politiques visant à produire, maintenir et utiliser pleinement le capital humain des individus. Dans cette définition, l'investissement social est défini par son objectif et interroge les politiques publiques à cet égard. Il constitue alors un référentiel critique pour les politiques françaises en interrogeant la pertinence des politiques pour répondre à cet objectif global, la cohérence entre les objectifs spécifiques aux différents champs, les moyens mis en œuvre et les objectifs opérationnels...

Les pratiques d'évaluation des politiques publiques à *la française* (Perret, 2008) peuvent alors être mobilisées pour questionner les politiques françaises à l'aune de l'investissement social et être davantage utilisées en amont des réformes futures.

procéder en deux étapes de calculs : une estimation de l'impact des politiques d'investissement social puis une analyse des coûts pour produire, suivant le besoin, des analyses coûts-efficacité ou coûts-bénéfices.

3.2. Le recours aux modèles théoriques, calibrations et simulations

Les analyses macroéconomiques des effets de l'investissement social peuvent aussi faire appel à des modèles théoriques, des simulations ou calibrations de modèles d'équilibre général calculables. Ces analyses s'inscrivent dans des cadres conceptuels particuliers, nécessitent un nombre important d'hypothèses et sont souvent très sensibles au choix des paramètres. Par exemple, Böhm, Grossmann, & Steger (2015) construisent un modèle d'équilibre général dynamique pour estimer si, théoriquement, l'expansion de l'éducation supérieure induit une croissance qui bénéficie à tous (*trickle-down economy*). Ils calibrent les paramètres de leur modèle à partir d'observations sur données américaines et de paramètres conventionnels de taux de croissance, d'inflation, etc. Ils concluent que promouvoir l'éducation supérieure implique d'abord une baisse de consommation et de revenu pour les personnes peu qualifiées mais qu'à l'issue d'une phase de transition (d'environ 50 ans...), ils en bénéficient également. En revanche, leur modèle prédit que les politiques qui se focalisent sur l'amélioration des compétences des travailleurs peu qualifiés produisent des effets contrastés en raison d'effets d'équilibre général défavorables. Les auteurs concluent donc que le mieux pour les travailleurs peu qualifiés reste la redistribution.

En analysant ces résultats à l'aune de l'investissement social, l'amélioration du capital humain des travailleurs peu qualifiés produit des effets négatifs sur leurs revenus pendant au moins 50 ans, ce qui ne semble pas vraiment souhaitable et rejoint les critiques de Cantillon (2011) sur l'effet Matthieu de l'investissement social.

Ces conclusions sont conditionnelles à la croyance que l'économie américaine peut être représentée par ce système d'équations, représentant des marchés en concurrence parfaites et que les valeurs des paramètres utilisés sont les bonnes. La temporalité de la simulation est bien sûr très discutable vu la vitesse du changement technologique des 30 dernières années. Le modèle considère en effet une élasticité de substitution capital-travail constante sur la période.

Au-delà de la discussion des hypothèses des modèles utilisés, une difficulté importante doit être soulignée. Dans de très nombreux cas (et notamment pour de nombreuses interventions que nous avons présentées plus tôt), il n'y a pas de modèle capable de prédire *ex-ante* l'impact ou les retombées d'une politique d'investissement sociale donnée ou, *a contrario*, il en existe plusieurs concurrents qui correspondent aux faits stylisés mais dont les implications ne sont pas du tout les mêmes. Un exemple extrêmement parlant de cette situation est à trouver dans la littérature sur l'éducation et la croissance où s'opposent notamment les modèles classiques ayant enrichi les travaux de Solow (Mankiw, Romer & Weil (QJE 1992) et les modèles de croissance endogène (Romer, 1986) (Lucas, 1988). Dans ces deux modèles, le capital humain joue un rôle positif sur la

croissance mais de façon très différente. Dans le premier cas, le capital humain entre comme facteur de production et son processus d'accumulation affecte la croissance. Dans le second cas, c'est au contraire le niveau de capital humain de l'économie qui affecte la croissance car il n'y a pas de rendement décroissant. Confrontés aux données, ces deux modèles fournissent des prédictions relativement cohérentes. Cependant, pour le premier modèle, un État qui investirait dans le capital humain de sa population (via une réforme éducative notamment) n'affecterait que de façon transitoire le taux d'accumulation du capital humain et n'aurait donc pas d'effet sur la croissance à long terme. Pour le second modèle, une augmentation du niveau d'éducation de la population placerait l'économie sur un nouveau sentier de croissance plus élevé. La littérature empirique testant ces modèles n'a pas permis de conclure à la domination d'un modèle sur l'autre (Durlauf, 2009).

3.3. Recenser les évaluations d'impact existantes

Des évaluations de l'impact de politiques d'investissement social existent déjà, pour certains champs en quantité significative, et doivent être collectées et synthétisées pour documenter les opportunités de nouvelles politiques ou expérimentations. Ces évaluations doivent être analysées de façon critique tant sur leur méthodologie que sur le contexte dans lesquelles elles ont été menées et dans quel environnement elles se situent. Ces informations sont cruciales pour anticiper la reproductibilité des résultats dans d'autres localités, etc.

La collecte et l'agrégation des preuves disponibles de l'efficacité de politiques d'investissement social est nécessaire à la prise de décision futur. Cependant, nous avons beaucoup discuté de la qualité des recherches « premières » – les expérimentations, quasi-expérimentations, etc. – mais la qualité des recherches secondaires est également primordiale.

Les revues de littérature ne constituent pas un exercice particulièrement périlleux et de nombreuses institutions, associations, organisations non gouvernementales... en produisent régulièrement sur les thématiques qui les concernent. Le danger de ces travaux est de tomber dans le « *cherry picking* » : consciemment ou inconsciemment⁴⁰ choisir les éléments de preuves qui renforcent les conclusions préférées. Ce type de travail est donc souvent (à) considéré(er) avec prudence. Une façon d'éviter ce problème est d'avoir recours à des revues systématiques. Ces approches visent l'exhaustivité et recensent autant de recherches pertinentes pour mesurer l'efficacité de l'investissement social pour une politique donnée. Elles utilisent pour cela des méthodes explicites d'identification des sources et des informations qu'elles contiennent pour pouvoir conclure. Elles permettent de scruter attentivement la qualité des résultats sur la base des projets de recherches et des méthodes employées. L'avantage de ces

⁴⁰ Il existe de nombreux biais cognitifs que l'expertise dans un domaine ne permet pas d'éviter (croire le contraire est en soit un biais dit métacognitif). En particulier le biais de confirmation qui consiste à valoriser davantage les éléments qui concourent à notre argument et négliger – inconsciemment ou no – les autres.

méthodes tient également au caractère explicite du processus de collecte des travaux, des critères d'inclusion et d'exclusion permettant en principe de répliquer les analyses.

Lorsque c'est possible, les résultats peuvent être agrégés dans de nouveaux indicateurs statistiques dans une méta-analyse.

L'intérêt est double : d'abord, on mobilise, pour une politique donnée, l'ensemble des informations disponibles sur son efficacité, indépendamment de son contexte, des modalités de mise en œuvre, etc. Ensuite, on identifie des paramètres d'efficacité moyenne de la politique et d'incertitude des résultats. On peut alors connaître l'effet moyen d'une politique et sa dispersion, au-delà de la variabilité des dispositifs et des pays et contextes dans lesquels ils sont mis en place. Ces éléments peuvent alors être utilisés dans d'autres analyses et notamment dans les calculs de rendement.

De plus en plus, ces méta-analyses tiennent compte de la qualité du design de la recherche et de la possibilité d'interpréter les résultats de façon causale. Certaines revues n'exploitent que des résultats d'expériences aléatoires contrôlées, d'autres sont un peu plus flexibles mais toutes admettent des critères de sélection rigoureux et explicites visant à assurer une bonne validité interne des travaux utilisés. Ces revues peuvent permettre d'évaluer avec un niveau de preuve élevé l'efficacité d'interventions spécifiques et, le cas échéant, encourager l'engagement de réformes investissement social fondé sur ces preuves *ex-ante* et intégrant un dispositif d'évaluation pour s'assurer des effets nouvellement produits.

Ainsi, la revue systématique de l'effet « d'école de lecture l'été » sur les performances d'élèves de la maternelle à un niveau équivalent à la 4^e de Kim & Quinn (2013) indique des effets globalement positifs sur diverses mesures de performances de lecture. Cette revue n'exploite que des travaux sur données américaines ce qui interroge la validité de ses conclusions dans un autre contexte. D'un autre côté, elle exploite 41 analyses indiquant que les cours d'été en classe avec un enseignant de littérature et les interventions à la maison basées sur des « activités de lecture entreprises par les enfants » améliorent significativement les performances en lecture. Les interventions qui se basent sur des méthodes d'enseignements fondées sur la recherche et qui incluent une majorité d'enfants d'origines sociales défavorisées ont des effets plus importants que des groupes plus mixtes. Ce type d'intervention apparaît donc prometteur pour remédier aux difficultés de lecture des élèves les plus en difficulté, certainement aussi en France. Cette intervention peut être complétée d'un transfert de revenus conditionnel afin d'inciter davantage les enfants de familles défavorisées à y souscrire. En effet, la méta-analyse de Baird, Ferreira, Özler, & Woolcock (2013) indique que les transferts monétaires aux familles améliorent significativement le taux d'inscription des enfants, et ce d'autant plus si ces transferts sont conditionnels et qu'il existe un mécanisme de contrôle. Cette étude est basée sur 35 travaux concernant les pays en développement.

Des institutions relativement récentes ont émergé pour élaborer ou financer des revues systématiques d'impact de politiques ou interventions dans différents champs. En particulier, Campbell Collaboration produit des revues de littérature systématiques dans

quatre domaines liés : la criminalité et la justice, l'éducation, le développement international et les politiques sociales. Dans leurs quelque 200 revues systématiques, on trouve, par exemple, une méta-analyse de l'efficacité des programmes de prévention du décrochage scolaire à partir de 152 études. En agrégeant les résultats de ces évaluations, ils concluent que les programmes de lutte contre le décrochage diminuent en moyenne de 70 % le risque de décrochage, avec un intervalle de confiance à 95 % compris entre 56 % et 90 %. Les auteurs identifient une douzaine de catégories d'interventions pour lesquels ils peuvent présenter un effet moyen et un intervalle de confiance.

Des travaux similaires existent sur l'efficacité des politiques d'activation sur le marché du travail, de programmes de soutien parental sur le développement des enfants, etc.

Les méta-analyses existantes ne sont pas toujours adaptées aux objectifs de l'investissement social même si elles analysent des politiques dont on peut attendre des effets en matière d'amélioration et de bonne utilisation du capital humain. Les réponses qu'elles fournissent peuvent être frustrantes de ce point de vue. Par exemple, les revues de littérature sur l'activation que nous avons recensées peinent à distinguer les effets sur la pauvreté, sur la qualité des emplois, etc.

Ainsi, d'autres recensions systématiques d'évaluations de politiques d'investissement social devront être produites en utilisant probablement des critères de sélection et des variables différentes. L'objectif est d'augmenter le niveau de preuve en cumulant plusieurs évaluations sur des résultats spécifiques à l'investissement social.

3.4. Les données inexploitées

Les analyses existantes et les revues systématiques ne permettent pas (encore) d'identifier l'impact de politiques d'investissement social dans tous les domaines qui y ont trait. Des analyses originales doivent donc être menées pour apporter de nouvelles pierres à l'édifice.

Nous ne parlons pas ici de la mise en œuvre de nouvelles politiques d'investissement social mais de celles existantes pour lesquelles il existe des ambiguïtés sur les effets induits, les interactions avec d'autres politiques, etc. Des réformes récentes des politiques familiales – pour lesquelles aucun protocole d'évaluation d'impact n'a pas été mis en œuvre à notre connaissance – sont de bons exemples de politiques qui peuvent améliorer le capital humain des jeunes enfants et la bonne utilisation de celui de leurs mères mais pour lesquelles il n'est aujourd'hui pas possible de répondre à la question de leurs effets. Cependant leur mise en œuvre et leurs règles d'éligibilité peuvent générer des situations quasi expérimentales potentiellement sources d'identification. Ainsi, la prestation partagée d'éducation de l'enfant n'est ouverte qu'aux parents d'enfants nés après le 1^{er} janvier 2014 ; aux parents qui ont cotisé au moins 8 trimestres, etc.

Il n'est pas du tout garanti que ces situations permettent une analyse quasi expérimentale satisfaisante⁴¹, mais nous plaidons pour une exploitation plus systématique des différences locales, des changements de législation, des critères d'éligibilité, des déploiements progressifs, pour produire des évaluations d'impact de l'investissement social.

Ce genre d'analyse n'est pas toujours possible en raison des informations disponibles, du manque d'accès des chercheurs aux données et de l'impossibilité d'apparier différentes sources. Ce problème n'est pas spécifique à l'investissement social, bien entendu, mais il se manifeste particulièrement dans ce champ où de nombreux producteurs de données ne relèvent pas de la statistique publique⁴² et ne peuvent donc les partager à des entités extérieures. Cette situation doit être modifiée par le projet de loi sur le numérique (dont la dernière version est celle modifiée par le Sénat en 3 mai 2016) mais les modalités de diffusion ne sont pas encore définies.

L'accès à certaines données administratives est possible en France dans des conditions très limitées ce qui garantit l'anonymat et le secret statistiques de ces ressources précieuses mais limite considérablement les recherches.

Ces fichiers ont une structure longitudinale – par nature – qui permet de suivre les individus dans le temps et ainsi de répondre à des questions cruciales dans le champ de l'investissement social tel que l'effet à long terme d'une perte d'emploi (Wachter, Schmieder, & Bender, 2009), l'effet d'une perte d'emploi sur la mortalité (Wachter & Sullivan, 2009), sur les inégalités et la mobilité salariale (Kopczuk, Saez, & Song, 2010), etc.

En dehors des États-Unis, de nombreux pays ont développé des systèmes d'accès aux données administratives pour les chercheurs. Au Danemark, Statistics Denmark prépare des données anonymes en combinant les informations de données administratives pour des projets de recherche approuvés suite à des appels d'offres concurrentiels. Les chercheurs accèdent alors aux données à distance à partir d'un serveur sécurisé (Card, Chetty, Feldstein, & Saez, 2010). C'est notamment par ce biais que Kleven, Landais & Søggaard (2015) ont pu étudier avec autant de précision l'effet de l'arrivée d'un enfant sur les trajectoires d'emploi et de salaires des hommes et des femmes.

⁴¹ Les nouvelles politiques ont tendance à être mal connues à leurs débuts ce qui peut limiter la validité de la comparaison des enfants autour du 1^{er} janvier 2014. Le fait d'avoir cotisé peu de trimestres peut, suivant l'âge, signaler un éloignement de l'emploi et la comparaison, même avec ceux qui ont juste 8 trimestres, peut être trop locale pour pouvoir caractériser l'efficacité du dispositif.

⁴² La loi sur la statistique publique de 1951 a été modifiée par la loi de modernisation de l'économie de 2008 qui élargit leur champ à « l'ensemble des exploitations de données administratives effectuées à des fins d'information générale hors du Service Statistique Public (Insee et services statistiques ministériels) à condition qu'elles le soient dans le cadre d'une mission de service public ».

3.5. Utiliser l'expérimentation

L'investissement social est aussi un chemin de réforme où les nouvelles politiques doivent démontrer leur effet.

La situation actuelle des finances publiques offre des opportunités intéressantes pour l'expérimentation. Les contraintes budgétaires induisent une rationalisation des ressources sur des bases territoriales ou démographiques pouvant créer des groupes de contrôles naturels dans certaines conditions. Mieux encore, certains programmes peuvent être progressivement déployés (exemple de la réforme des rythmes scolaires de 2013) et fournir des terrains propices à des évaluations expérimentales ou quasi expérimentales. Il importe bien sûr que l'évaluation soit initialement prévue pour permettre la collecte des informations nécessaires et s'assurer de la comparabilité des groupes affectés et non affectés (ce qui n'a pas été le cas de la réforme des rythmes). La plus grande autonomie donnée aux collectivités locales offre également des opportunités d'expérimentations intéressantes.

Nelson & Sandberg (2016) plaident aussi pour l'utilisation des expérimentations sociales dans la définition des politiques d'investissement social en analysant la mise en œuvre de transferts monétaires conditionnels en Amérique Latine. Les études à partir du programme Progressa, puis Esperanza et d'autres dispositifs similaires visant à transférer des ressources aux parents pauvres à condition de scolariser les enfants sont nombreuses et concluent presque systématiquement à des effets positifs de ces dispositifs (Kabeer & Waddington, 2015). Pour Nelson et Sandberg, les politiques d'investissement social universelles peuvent avoir tendance à laisser de côté les populations pauvres et défavorisées (l'effet Matthieu décrit notamment par (Cantillon & Lancker, 2011) (Cantillon & Van Mechelen, 2013)) et alors, des politiques de transferts conditionnels semblent adaptées pour y remédier. Les auteurs concluent en outre que les politiques d'investissement social doivent interroger la quantité d'investissement nécessaire pour être efficaces, identifier les mécanismes causaux par lesquels transitent les effets et quelles conditions sont nécessaires pour que ces mécanismes soient effectifs.

Les méthodes expérimentales telles que les expérimentations aléatoires contrôlées ne sont pas adaptées à toutes les politiques d'investissement social. Nous ne prôtons pas une évaluation fondée sur une méthode particulière – où l'on ferait de l'évaluation là où c'est possible en délaissant les politiques où elles ne sont pas applicables. Les évaluations doivent être fondées sur la méthode la plus appropriée et elles seront souvent non expérimentales. Comme on doit interroger la pertinence des politiques d'investissement social, on doit opter pour les méthodes d'évaluation les plus pertinentes pour répondre aux questions de recherche qui se posent. Et ces questions se posent au cas par cas.

3.6. Identifier le rendement d'investissements sociaux.

Les possibilités d'identifier empiriquement le retour sur investissement restent faibles sur un plan « macro » pour les mêmes raisons qu'il est très difficile d'identifier l'effet macro des investissements sociaux. Les calculs de rendement d'investissements sociaux s'appliquent à des niveaux plus microéconomiques en employant les méthodes de calculs de coûts-efficacité et coûts-bénéfices.

Les analyses que nous avons présentées dans la section précédente exploitent le résultat de l'évaluation d'impact de programmes spécifiques et, *ex-post*, estiment l'efficacité du programme par des calculs coûts-efficacité ou coûts-bénéfices. Il s'agit bien d'une étape supplémentaire à l'identification de l'effet de l'investissement social et cela nécessite de nouvelles hypothèses pour lesquelles, aujourd'hui, il ne semble pas y avoir de consensus sur la méthode et les valeurs à utiliser.

Il semble nécessaire d'engager une réflexion globale sur l'usage du calcul socioéconomique appliqué aux politiques sociales afin d'établir un cadre conventionnel permettant au moins d'utiliser des valeurs tutélaires et prix implicites communs et au mieux, de fixer un référentiel de bonnes pratiques.

Lorsqu'on ne s'intéresse pas à un programme spécifique, qu'on souhaite estimer le rendement d'investissements sociaux dans un domaine, ou qu'on a besoin *ex-ante* de valoriser les avantages et les coûts d'un programme social prospectif, les informations nécessaires à ces réalisations ne sont généralement pas immédiatement disponibles ou directement exploitables.

Au-delà des problèmes de valorisation des coûts et d'intégration des résultats des évaluations dans les calculs, l'identification du rendement des investissements sociaux soulève plusieurs défis, au premier rang desquels figure l'intégration de la grande incertitude dans les prévisions et les valorisations qu'implique ce type d'analyse.

Il est fondamental que la pratique de répliques stochastiques des analyses coûts-bénéfices devienne la norme pour l'évaluation du rendement de l'investissement social afin de tenir compte de façon claire de l'incertitude qui pèse sur ces analyses.

Nous notons par ailleurs l'intérêt de comparer, à partir de calculs coûts-bénéfices, plusieurs réformes visant à répondre à un même objectif de politique publique. En appliquant les mêmes conventions, la comparaison permet de sélectionner les investissements les plus socialement rentables. Le fait de tester plusieurs alternatives rend plus acceptable pour le décideur le choix d'un autre projet que le sien si les preuves de l'efficacité de la politique alternative sont apportées de façon rigoureuse (Campbell, 1969). Ces comparaisons se basent sur des tailles d'effet obtenues à partir d'évaluations de projets comparables, solides et de préférence répliquées (ou issues de revues systématiques).

Conclusion

L'investissement social fait la promesse que la dépense sociale engagée aujourd'hui rapportera à un moment donné dans le futur si elle est orientée vers la dotation des individus en capital humain. Il ne peut avoir de traduction opérationnelle qu'à la condition – forte – d'être capable de démontrer ces retours.

Dans ce document, nous avons analysé la littérature consacrée aux effets de l'investissement social sur la société et questionné le choix des méthodes et du niveau d'analyse.

Les approches macroscopiques d'évaluation de l'effort d'investissement social ne peuvent se faire que sur une base conventionnelle (et non empirique), en comparant l'effort de la nation sur des « blocs » de politiques considérées conventionnellement comme relevant de l'investissement social... Les possibilités de « preuves » empiriques du retour sur investissement restent faibles.

Une évaluation utile des effets de l'investissement social suppose des niveaux de preuve suffisants, fondés sur une démarche d'exigence et de robustesse des résultats empiriques.

Estimer l'impact d'une politique d'investissement social implique de comparer la réalité avec la situation qui prévaudrait en l'absence d'investissement, donc de recourir à des évaluations d'impact utilisant des méthodes ou des contextes d'évaluations permettant d'identifier le contrefactuel. Dès lors, le niveau d'observation adapté n'est plus macro mais méso : chaque politique sociale, voire chaque dispositif d'une politique sociale. L'objectif de l'investissement social devient alors, pour une politique donnée, de maximiser son impact.

De telles évaluations existent déjà, pour certains champs en quantité significative. Elles doivent être collectées et synthétisées pour documenter les opportunités de nouvelles politiques ou expérimentations. Les revues systématiques apparaissent adaptées pour proposer des résultats synthétiques de la recherche, en tenant compte de la rigueur de l'évaluation et en limitant l'effet du contexte dans lequel la politique est mise en œuvre.

Prouver l'existence d'impacts n'est pas suffisant pour démontrer le rendement des investissements sociaux mais la question de l'efficience de la dépense sociale peut être traitée en intégrant les résultats des évaluations d'impacts dans des modèles de calculs socioéconomiques.

Cette dernière étape nécessite de résoudre de nombreux problèmes parmi lesquels figurent au premier rang :

- l'identification des coûts de mise en œuvre des politiques et des valeurs sociales des effets estimés ;

- la prise en compte de l'incertitude inhérente à l'évaluation des programmes sociaux dans l'estimation du rendement de l'investissement social.

Pour ce dernier point, nous avons proposé une solution générique utilisant des simulations de Monte Carlo.

En revanche, pour l'identification des coûts et des valeurs des effets sur la société, il semble nécessaire d'engager une réflexion globale sur l'usage du calcul socioéconomique appliqué aux politiques sociales afin d'établir un cadre conventionnel permettant au moins d'utiliser des valeurs tutélaires et prix implicites communs et au mieux, de fixer un référentiel de bonnes pratiques.

Une stratégie d'investissement social pour la France est une stratégie fondée sur l'efficacité des politiques publiques pour produire et maintenir le capital humain de la population (et permettre son bon usage). Elle s'intègre en ce sens plus largement dans le champ des politiques publiques fondées sur la preuve et suppose alors de s'intéresser davantage aux détails qu'aux politiques globales. L'apport de l'évaluation d'impact et des travaux les liant avec les coûts vise à hiérarchiser la pertinence de certains dispositifs et à soutenir les politiques sociales en démontrant leurs effets et leur rendement.

Bibliographie

- Ackerman, D. J. (2005). Getting teachers from here to there: Examining issues related to an early care and education teacher policy. *Early Childhood Research & Practice*, 7(1).
- Adam, S., Bozio, A., & Emmerson, C. (2009). *Can we estimate the impact of the Choices package in Pathways to Work ?* Department for Work and Pensions . Institute for Fiscal Study.
- Adam, S., Bozio, A., Emmerson, C., Greenberg, D., & Knight, G. (2008). *A cost-benefit analysis of Pathways to Work for new and repeat incapacity benefits claimants*. Department for Work and Pensions. London: Institute for Fiscal Study.
- Ahn, S.-H., & Kim, S.-W. (2015). Social investment, social service and the economic performance of welfare states. *International Journal of Social Welfare*, 24, 109-119.
- Algan, Y., Beasley, E., Vitaro, F., & Tremblay, R. E. (2016). *Impact of Non-Cognitive Skills Training on Academic and Non-academic Trajectories: From Childhood to Early Adulthood*. Working Paper.
- Alliance for useful evidence. (2016). *Using research evidences: a practical guide*. London: Nesta.
- Almond, D., & Currie, J. (2011). Human capital development before age five. In O. a. Ashenfelter, *Handbook of Economics of Education* (Vol. 4B, pp. 1315–1486.).
- Angrist, J. D., & Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review*, 88(3), 450-477.
- Angrist, J. D., & Krueger, A. B. (1991). Does compulsory schooling attendance affect schooling and earning ? *Quarterly Journal of Economics*, 106(4), 979-1014.
- Angrist, J., & Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114, 533–575.
- Böhm, S., Grossmann, V., & Steger, T. M. (2015). Does expansion of higher education lead to trickle-down growth? *Journal of Public Economics*, 132, 79-94.
- Bagshaw, S. M., & Bellomo, R. (2008). The need to reform our assessment of evidence from clinical trials: A commentary. *Philosophy, Ethics, and Humanities in Medicine*, 3(1), 23.
- Baird, S., Ferreira, F. H., Özler, B., & Woolcock, M. (2013). Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review. *Campbell Systematic Reviews*, 8.
- Baker, M., J., G., & Milligan, K. (2008). Universal child care, maternal labor supply, and family wellbeing. *Journal of Political Economy*, 116(4), 709-745.
- Banerjee, A. V., & Duflo, E. (2009). *The Experimental Approach to Development Economics*. Massachusetts Institute of technology,. Department of Economics and Abdul Latif Jameel Poverty Action Lab.
- Barnett, W. (1995). Long-term effects of early childhood programs on cognitive school outcomes. *Future of Children*, 5(3), 25-50.
- Bauernschuster, S., & Schlotter, M. (2015). Public child care and mothers' labor supply : Evidence from two quasi-experiments. *Journal of Public Economics*, 123, 1-16.
- Bauernschuster, S., & Schlotter, M. (2015). Public child care and mothers' labor supply Evidence from two quasi-experiments. *Journal of Public Economics*, 123, 1-16.
- Bayet, A. (1996). Carrières continues, carrières incomplètes et salaires. *Économie et Statistique*, 299, 21-36.
- Becker, E., Hayllar, O., & Wood, M. (2010). *Pathways to Work: programme engagement and work patterns. Findings from follow-up surveys of new and repeat and existing incapacity benefits customers in the Jobcentre Plus pilot and expansion areas*. Norwich: Department for Work and Pensions.

- Behaghel, L., Crépon, B., & Gurgand, M. (2014). Private and Public Provision of Counseling to Job-Seekers: Evidence from a Large Controlled Experiment. *Published in American Economic Journal: Applied Economics*, 6(4), 142-174.
- Benhenda, A., & Grenet, J. (2015). *Evaluation du coût du redoublement*. institut des politiques publiques.
- Benson, K., & Hartz, A. (2000). Comparison of observational studies and randomized, controlled trials. *The New England Journal of Medicine*, 342, 1878-86.
- Bernal, R., & Keane, M. P. (2011). Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers. *Journal of Labor Economics*, 29(3), 459-512.
- Bichot, J. (2010). *Le coût du crime et de la délinquance*. Paris: Institut pour la Justice et l'égalité des citoyens.
- Blanden, J., Bono, E. D., McNally, S., & Rabe, B. (2015, 05). Universal pre-school education: The case of public funding with private provision. *The economic journal*, 126, 682–723.
- Blatchford, P., Goldstein, H., Martin, C., & Browne, W. (2002). A study of class size effects in english school reception year classes. *British Educational Research Journal*, 28(2).
- Blatchford, P., Russell, A., & Webster, R. (2011). *Reassessing the Impact of Teaching Assistants : How Research Challenges Practice and Policy*. London: Routledge.
- Boisson-Cohen, M., & Collombet, C. (2016). *L'investissement social: quelle stratégie pour la France. Éléments pour le débat*. France Stratégie. Paris: France Stratégie.
- Bonoli, G. (2009). Varieties of social investment in labour market policy. In N. Morel, B. Palier, & J. Palme, *What Future for Social Investment?* (pp. 55-66). Stockholm: Institute for Futures Studies.
- Bonoli, G. (2012). Active labour market policy and social investment: a changing relationship. In N. Morel, B. Palier, & J. Palme, *Towards a Social Investment Welfare State? Ideas, Policies and Challenges* (pp. 181–204). Bristol: Policy Press.
- Bonoli, G. (2012). Comment on Anton Hemerijck. *Sociologica*, 1-5.
- Bontout, O., Maestri, V., & Vaalavuo, M. (2015). The efficiency and effectiveness of social protection systems over the life course. In E. Commission, *Employment and social developments in Europe 2015*. Bruxelles.
- Botheffled, S. (2016). L'approche allemande de l'investissement social. *L'investissement social : quelle stratégie pour la France*. Paris: France Stratégie.
- Bouguen, A. (2016). Adjusting content to individual student needs: Further evidence from an in-service teacher training program. *Economics of Education Review*, 50, 90-112.
- Bressoux, P., & Lima, L. (2011). La place de l'évaluation dans les politiques éducatives : le cas de la taille des classes à l'école primaire en France. *Raisons Éducatives*, 15, 99–123.
- Bressoux, P., & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris: PUF.
- Bressoux, P., Gurgand, M., Guyon, N., Monnet, M., & Pernaudet, J. (2016). *Evaluation des Programmes de Réussite Educative*. Institut des politiques publiques.
- Bressoux, P., Kramarz, F., & Prost, C. (2009). Teachers training, class size and students outcomes : learning from administrative forecasting mistakes. *The Economic Journal*, 119(536).
- Browning, M., & Heinesen, E. (2007). Class size, teacher hours and educational attainment. . *The Scandinavian Journal of Economics*, 109(2).
- Burgoon, B. (2015). *Practical Pluralism in the Empirical Study of Social Investment: Examples from Active Labor Market Policy*. Papier pour Hemerijck, ed. "Social Investment and its critics, University of Amsterdam.
- Caille, J. (2001, Juillet-Septembre). Scolarisation à 2 ans et réussite de la carrière scolaire au début de l'école élémentaire. *Éducation et Formations*, 60.
- Caille, J., & Rosenwald, F. (2006). Les inégalités de réussite à l'école élémentaire : construction et évolution. *France, Portrait social, dossier les inégalités de réussite*.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24(4), 409-429.

- Campbell, D. T. (1999). *The experimenting society*.
- Cantillon, B., & Lancker, W. V. (2011, 4). Le paradoxe de l'État d'investissement social: pourquoi la pauvreté n'a-t-elle pas baissé ? *Reflets et perspectives de la vie économique*, L, 55-73.
- Cantillon, B., & Van Mechelen, N. (2013). *Poverty reduction and social security: Cracks in a policy Paradigm*. Working Paper , CSB.
- Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grant, & R. Swidinsky, *Aspects of labour market behaviour: essays in honour of John Vanderkamp* (pp. 201-222). Toronto: University of Toronto Press.
- Card, D., Chetty, R., Feldstein, M., & Saez, E. (2010). *Expanding Access to Administrative Data for Research in the United States*. National Science Foundation .
- Card, D., Kluve, J., & Weber, A. (2010). Active labor market policy evaluations: a meta-analysis. *NBER Working paper*.
- Card, D., Kluve, J., & Weber, A. (2010). Active labor market policy evaluations : A meta-Analysis. *The economic Journal*, 120.
- Card, D., Kluve, J., & Weber, A. (2015, 7). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *IZA Discussion Paper* (9236).
- Cavaco, S., Fougère, D., & Pouget, J. (2009, 6). Estimating the Effect of a Retraining Program on the Re-Employment Rate of Displaced Workers. *IZA discussion paper*, 4227.
- Chauffaut, D., & Boyer, D. (2010). Retombées économiques des politiques d'accueil de la petite enfance: Quelques approches récentes. *Politiques sociales et familiales*, 99, 89-96.
- Chesné, J. F. (2014). *D'une évaluation à l'autre : des acquis des élèves sur les nombres en sixième à l'élaboration et à l'analyse d'une formation d'enseignants centrée sur le calcul mental*. Paris: Université Paris 7 – Denis Diderot.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2010). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR,. *NBER Working Papers* 16381.
- Commission Européenne. (2013). *Investir dans le domaine social en faveur de la croissance et de la cohésion, notamment par l'intermédiaire du Fonds social européen, au cours de la période 2014-2020*. Bruxelles: COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS.
- Commission Européenne. (2015). *Policy Roadmap for the implementation of the Social Investment Package*. European Commission.
- Cook, P. J., Dodge, K., Farkas, G., Roland G. Fryer, J., Guryan, J., Ludwig, J., . . . Steinberg, L. (2014). *The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago*. NBER.
- Crawford, C., Dearden, L., & Greaves, E. (2013). *When you are born matters : evidence for England*. Institute for Fiscal Study.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., & Zamora, P. (2013). Do Labor Market Policies Have Displacement Effects: Evidence from a Clustered Randomized Experiment. *Quarterly Journal of Economics*, 128(2), 531-580.
- Crépon, B., Ferracci, M., & Fougere, D. (2012). Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence? . *Annales d'Economie et de Statistique*, 107-108, 175-199.
- Crepon, B., Gurgand, M., kamionka, T., & Lequien, L. (2012). *Is Counseling Welfare Recipients Cost-Effective? Lessons from a Random Experiment*. Working Paper.
- Dalziel, K. M., Halliday, D., & Segal, L. (2015). Assessment of the Cost–Benefit Literature on Early Childhood Education for Vulnerable Children: What the Findings Mean for Policy. *SAGE Open*(1-3), 1-14.
- Damond, J. (2015). L'investissement social : contenu et portée d'une notion en vogue. *Revue de droit sanitaire et social*, 4, 722-733.

- Dang, H.-A., & Rogers, F. H. (2008). *How to interpret the growing phenomenon of private tutoring : human capital deepening, inequality increasing, or waste of resources ?* The World Bank.
- DARES. (2016). *Rapport intermédiaire*. Paris: Comité scientifique en charge de l'évaluation de la Garantie Jeunes.
- Datar, A. (2006). (2006). Does delaying kindergarten entrance give children a head start ? . *Economics of Education review*, 25, 43-62.
- Datta Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality childcare. *Journal of Public Economics*, 94(1-2), 30-40.
- De Deken, J. (2013). Identifying the skeleton of the social investment state: defining and measuring patterns of social policy change on the basis of expenditure data. In B. Cantillon, & F. Vandenbroucke, *For better, for worse. For richer, for poorer. Labour Market Participation, Social Redistribution and Poverty in the EU*. Oxford University Press.
- De Deken, J. (2016). Conceptualiser et mesurer l'investissement social dans une perspective comparative. *Information sociale*, 192, 28-38.
- Delors, J., & Dollé, M. (2009). *Manifeste pour une politique sociale. Investir dans le social*. Odile Jacob.
- Desplatz, R., & Ferracci, M. (2016). *Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens*. Paris: France Stratégie.
- Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2012). *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education*. Cambridge MA: Abdul Latif Jameel Poverty Action Lab (J-PAL), MIT.
- Duflo, E. (2005, 2). Évaluer l'impact des programmes d'aide au développement : le rôle des évaluations par assignation aléatoire. *Revue d'économie du développement*, 13, 185-226.
- Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 5, 1739-74.
- Dumas, C., & Lefranc, A. (2010). *Early schooling and later outcomes : Evidence from pre-school extension in France*. Cergy: THEMA (Théorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.
- Durlauf, S. N. (2009). The Rise and Fall of Cross- Country Growth Regressions. *History of Political Economy*, 41.
- Duvoux, N. (2016). Lutter contre la pauvreté: coût social ou investissement. *L'investissement social: quelle stratégie pour la France : séance pauvreté*. Paris: Apprentis d'auteuil - Cnaf - DGCS - France Stratégie - Science Po.
- Ecalte, J., Labat, H., Le Cam, M., Rocher, T., Cros, L., & Magnant, A. (2015, 08). Evidence-based practices to stimulate emergent literacy skills in kindergarten in France: A large-scale study. *Teaching and Teacher Education*.
- Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). Early childhood education. *NBER Working paper*.
- Elbaum, M. (2010). *La politique familiale : des « retours sur investissement » qui dépendent de choix économiques et sociaux*. Paris: OFCE.
- Esping-Andersen, G. (2002). *Why We Need a New Welfare State*. Oxford University Press.
- Esping-Andersen, G. (2009). *Incomplete Revolution: Adapting Welfare States to Women's New Roles*. Polity.
- Esping-Andersen, G., & Pallier, B. (2008). *Trois leçons sur l'Etat providence*. Paris: Seuil, collection La République des idées.
- ESSEC. (2011). *Guide du retour social sur investissement (SROI)*. Cergy: Institut de l'Innovation et de l'Entrepreneuriat social.
- Evertson, C., & Randolph, C. (1989). Teaching practices and class size : a new look at an old issue. *Peabody Journal of Education*, 67.

- Fack, G., & Grenet, J. (2015). Improving College Access and Success for Low-Income Students: Evidence from a Large Need-based Grant Program. *American Economic Journal: Applied Economics*, 7(2), 1-32.
- Felfe, C., & Lalive, R. (2014). Does early child care help or hinder child development? *IZA Discussion Paper*, 8484.
- Fertig, M., & Kluve, J. (2005). *The effect of age at school entry on educational attainment in germany*. IZA.
- Filatriau, O., Fougère, D., & Tô, M. (2013). *Will sooner be better ? The impact of early preschool enrollment on cognitive and non-cognitive achievement of children*. center for economic policy research.
- Filges, T., Smedslund, G., Knudsen, A.-S. D., & Jørgensen, A.-M. K. (2015). Active Labour Market Programme Participation for Unemployment Insurance Recipients: A Systematic Review . *Campbell Systematic Reviews*, 2.
- Fitzpatrick, M. (2008). Starting school at four: the effect of universal pre-kindergarten on children's academic achievement. *Journal of Economic Analysis and Policy*, 8(1), 1-40.
- Florin, A. (2007). *L'école primaire en France*. Paris: Haut Conseil de l'Éducation.
- Florin, A., Guimard, P., & et Nocus, I. (2014). *Evaluation du dispositif Coup de pouce clé (2013-2014)*. . Rapport pour l'association Coup de pouce clé.
- Fredriksson, P., Öckert, B., & Oosterbeek, H. (2013). Long-term effects of class size. . *The Quarterly Journal of Economics* , 128(1).
- Fryer, R. G. (2016). The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments.
- García, J. L., Hojman, A., & Shea, J. (2016, Janvier). The Opportunity Cost of Early Childhood Education: Formal, Informal, and Maternal Care. *Manuscrit non achevé* .
- Gary-Bobo, R., & Mahjoub, M.-B. (2013). Estimation of class-size effects, using maimonides rule and other instruments : the case of french junior high schools. . *Annals of Economics and Statistics*, 111, 193–225.
- Givord, P., & Marbot, C. (2015). Does the cost of child care affect female labor market participation? An evaluation of a French reform of childcare subsidies. *Labour Economics*, 36, 99-111.
- Gormley, W., & Gayer, T. (2005). Promoting school readiness in Oklahoma. *Journal of Human Resources*, 40(3), 553-558.
- Gormley, W., & T., G. (2005). Promoting school readiness in Oklahoma. *Journal of Human Resources*, 40(3), 553-558.
- Goujard, A., & L'Horty, Y. (2009). *La définition des zones témoin pour l'expérimentation du Revenu de Solidarité Active*. Centre d'Études des Politiques Économiques (EPEE), . Université d'Evry Val d'Essonne.
- Goux, D., & Maurin, E. (2005). The effect of overcrowded housing on children's performance at school. *Journal of Public Economics*, 89, 797– 819.
- Goux, D., & Maurin, E. (2010). Public school availability for two-year olds and mothers' labour supply. *Labour Economics*, 17, 951-962.
- Goux, D., Gurgand, M., & Maurin, E. (2013). *Evaluation d'impact du dispositif Coup de pouce clé*. Rapport au Fonds d'expérimentation pour la jeunesse, Paris.
- Gravel, N., Hagneré, C., Picard, N., & Trannoy, A. (2001). Une évaluation de l'impact incitatif et redistributif d'une réforme des minima sociaux. *Revue Française d'Économie, Programme National Persée*, 16(1), 125-167.
- Graves, J. (2013). School calendars, child care availability and maternal employment. *Journal of Urban Economics*, 78(75-70).
- Grenet, J. (2010). *Academic performance, Educational Trajectories and the Persistence of Date of Birth Effects. Evidence from France*. Paris: Paris School of Economics.

- Gurgand, M., & Montialoux, C. (2009). *Impact du RSA sur les taux d'entrée en emploi des bénéficiaires du RMI*. note préparée pour le rapport du Comité national d'évaluation du RSA.
- Hall, P. (1993). Policy paradigms, social learning and the state: the case of economic policymaking in Britain. *Comparative Politics*, 25(3), 275-296.
- Havnes, Y., & Mogstad, M. (2011, 4). No child left behind: subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, 3, 97-129.
- Heckman, J. (2000). Policies to foster human capital. *Research in economics*, 54(1), 3-56.
- Heckman, J. J. (1992). Randomization and social policy evaluation. In C. Manski, & I. Garfinkel, *Evaluating Welfare and Training Programs*. Cambridge, MA: Harvard University Press.
- Heckman, J. J., LaLonde, R. J., & Smith, J. A. (1999). The Economics and Econometrics of Active Labor Market Programs. In O. Ashenfelter, & D. Card, *Handbook of labor economics* (Vol. 3). Amsterdam: Elsevier.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. Q. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), 114-128.
- Heckman, J., Moon, S., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative economics*, 1(1), 1-46.
- Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6), 2052-2086.
- Heckman, J., S., M., Pinto, R., Savelyev, P., & Yavis, A. (2010). The rate of return to the High Scope Perry pre-school Program. *Journal of Public Economics*, 94(1-2), 114-28.
- Heim, A., & Ni, J. (2016). *L'éducation peut-elle favoriser la croissance ?* Paris: France Stratégie.
- Heim, A., Seinmetz, C., & Tricot, A. (2015). *Faut-il encore redoubler ? Conférence de consensus*. Paris: Canopé.
- Hemerijck, A. (2012). Two or three waves of welfare state transformation? In N. Morel, B. Palier, & J. Palme, *In Towards a Social Investment Welfare State? Ideas, Policies and Challenges* (pp. 33-60). Bristol: Policy Press.
- Hemerijck, A. (2013). *Changing Welfare States*. Oxford University Press.
- Hemerijck, A. (2014). Social investment "stocks", "Flows" and "Buffers". *Politiche Sociali*, 1, 9-26.
- Hemerijck, A., & Vydra, S. (2016). Le champ d'analyse de la politique d'investissement social. *Information social*, 192, 10-20.
- Huber, E., & Stephens, J. D. (2012). Post industrial social policy. *The Future of Democratic Capitalism*. Duke University.
- Imbens, G. W., & Wooldridge, J. M. (2008). *Recent developments in the econometrics of program evaluation*. IZA.
- Jenson, J. (2012). Redesigning Citizenship Regimes after Neoliberalism. Moving Towards Social Investment. In *Towards a Social Investment Welfare State? Ideas, Policies and Challenges*. Nathalie Morel ; Bruno Palier ; Joakim Palme.
- Kabeer, N., & Waddington, H. (2015). Economic impacts of conditional cash transfer programmes: a systematic review and meta-analysis. *Journal of Development Effectiveness*, 7(3), 290-303.
- Karoly, L., Kilburn, R., & Cannon, J. (2005). *Early Childhood Interventions: Proven Results, Future Promise*. Santa Monica, CA: RAND Corporation.
- Keslair, F., Maurin, E., & McNally, S. (2012). Every child matters? An evaluation of "Special Educational Needs" programmes in England. *Economics of Education Review*, 31(6), 932-948.
- Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from Kindergarten to Grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, 83(3), 386-431.

- Kleven, H. J., Landais, C., & Sjøgaard, J. E. (2015). *Children and Gender Inequality: Evidence from Denmark*. London: LSE.
- Knight, G., Salis, S., Francavilla, F., Radu, D., Hevenstone, D., Mocca, E., & Tousley, B. (2013). *Provider-led Pathways to Work Net impacts on employment and benefits*. London: Department for Work and Pensions.
- Kopczuk, W., Saez, E., & Song, J. (2010). Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937. *Quarterly Journal of Economics*, 125(1), 91-128.
- Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics*.
- Kvist, J. (2016). *Social investment Reforms: How to achieve more with less*. Roskilde University, Institute of Society and Globalization. Science Po Seminar.
- Laferrère, A., & Le Blanc, D. (2002). Comment les aides au logement affectent-elles les loyers ? *Économie et Statistique*, 351(1), 3-30.
- Lancker, W. V. (2013). Putting the child-centred investment strategy to the test: evidence for the EU27. *European Journal of Social Security, Volume 15 (2013), No. 1, 15(1)*.
- Lazear, E. P. (2001, 8). Educational Production. *The Quarterly Journal of Economics*, 116(3), 777-803.
- Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281-355.
- Lefebvre, P., Merrigan, P., & Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a Canadian natural experiment on low-fee universal child care. *Labor Economics*, 16, 490-502.
- Levin, H. M., & McEwan, P. J. (2001). *Cost-Effectiveness Analysis*. (S. P. Inc., Ed.) CA:: Thousand Oaks, 2nd Edition.
- L'Horty, Y., & Petit, P. (2010). *Evaluation aléatoire et expérimentations sociales*. EPEE. Centre d'étude des politiques économiques de l'université d'Evry.
- Lochner, L. (2011). Non-Production Benefits of Education: Crime, Health, and Good Citizenship . In E. Hanushek, S. J. Machin, & L. Woessmann, *Handbook of Economics of Education* (Vol. 4).
- Loncle, P., Muniglia, V., & Rivard, T. (2009). *La mise en oeuvre de l'expérimentation du RSA: Synthèse réalisée à partir d'enquêtes qualitatives dans cinq départements français*. École des Hautes Etudes en Santé Publique (EHESP), département ShScS , Laboratoire étude, recherche, formation, en action sociale (LERFAS). Comité d'Évaluation des expérimentations.
- Lucas, R. J. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22, 3-42.
- Mühlenweg, A., Blomeyer, D., Stichnotha, H., & Laucht, M. (2012). Effects of age at school entry (ase) on the development of non-cognitive skills : Evidence from psychometric data. *Economics of Education review*, 31, 68-76.
- Manning, M., Garvis, S., Fleming, C., & Wong, G. T. (2015, 3). Protocol The Relationship between teacher qualification and the quality of the early childhood care and learning environment: a systematic review. *Campbell collaboration*.
- Marston, G., & Watts, R. (2003). Tampering with the evidence: A critical appraisal of evidence-based policymaking. *The Drawing Board: An Australian Review of Public Affairs*, 3(3), 143-163.
- Maurin, E., & Roy, D. (2008). *L'effet de l'obtention d'une place en crèche sur le retour à l'emploi des mères et leur perception du développement de leurs enfants*. Paris: CEPREMAP.
- McEwan, P. J., & Shapiro, J. S. (2008). The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates. *The journal of human resources*, 43, 1-29.
- McNally, S., & Machin, S. (2008). The literacy hour. *Journal of Public Economics*, 92, 1441-1462.
- Midgley, J. (1999). Growth, redistribution, and welfare: Toward social investment. *The Social Service Review*, 73(1).

- Miguel, E., & Kremer, M. (2004, 1). Worms: Identifying Impacts on Education and Health in the Presence of Externalities. *Econometrica*.
- Miller, T. R., Cohen, M. A., & Wiersema, B. (1996). *Victim costs and consequences: a new look*. Washington D.C.: National Institute of Justice.
- Morel, N., Palier, B., & Palme, J. (2013). *The Social Investment Welfare State in Europe, 1990s and 2000s: Economic Ideas and Social Policies*. Science Po. Governance and Globalization Working Paper Series.
- Morgan, K. J. (2012). Promoting social investment through work-family policies: which nations do it and why? In *Towards a Social Investment Welfare State? Ideas, Policies and Challenges*. Nathalie Morel; Bruno Palier ; Joakim Palme.
- Moschion, J. (2007). *Reconciling work and family life: the effect of French family policies*. Centre d'Economie de la Sorbonne.
- Nelson, M. (2008). *Education policy and the consequences for Labor market integration in denmark, Germany, and the netherlands*. PhD Thesis, Chapel Hill.
- Nelson, M., & Sandberg, J. (2016, 2). From perspectives to policy contingencies: Conditional cash transfers as social investments. *Global social policy*.
- Nelson, M., & Stephens, J. D. (2009). Human Capital Policies and the Social Investment Perspective: Explaining the Past and Anticipating the Future. In *What Future for Social Investment?* Nathalie Morel ; Bruno Palier ; Joakim Palme.
- Nikolai, R. (2012). Towards Social Investment? Patterns of Public Policy in the OECD World. In B. P. Press. (Ed.), *Towards a Social Investment State? Ideas, Policies and Challenges*. . N. Morel ; B. Palier ; J. Palme.
- Nolan, B. (2013). What use is 'social investment' ? *Journal of European Social Policy* , 23, 459.
- Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world : (under) investing in the very young. *Economics of Education review*, 29, 271-282.
- Olsen, S. (2003). *Social Return on Investment: Standard Guidelines*. Center for Responsible Business UC Berkeley. Escholarship.
- Paetzold, J., & Vliet, O. V. (2014, 9). EU Co-Ordination and the Convergence of Domestic Unemployment Protection Schemes. *Journal of Common Market Studies*, 52, 1070–1089.
- Palier, B. (2005). Vers un État d'investissement social. Pistes pour une redéfinition de la protection sociale. *Informations sociales*, 8(128), 118-128.
- Palier, B. (2014). *La stratégie d'investissement social*. CESE, Paris.
- Parent, G., Sautory, O., & Desplatz, R. (2013). *L'accompagnement des demandeurs d'emploi : enseignements des évaluations*. DARES, direction de l'animation de la recherche, des études et des statistiques. Paris: DARES.
- Paul, J.-J., & Troncin, T. (2004). *Les apports de la recherche sur l'impact du redoublement comme moyen de traiter les difficultés scolaires au cours de la scolarité obligatoire*. Haut conseil de l'évaluation de l'école.
- Perret, B. (2007). Evaluation et politique du chiffre. *Esprit*, 212.
- Perret, B. (2008). *L'évaluation des politiques publiques*. Paris: Repères - La découverte.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses . *Journal of Epidemiology and Community Health*, 57(7), 473.
- Piketty, T. (1998). L'impact des incitations financières au travail sur les comportements individuels : une estimation pour le cas français. *Économie et Prévision*, 132(1), 1-35. .
- Piketty, T. (2003). *L'impact de l'allocation parentale d'éducation sur l'activité féminine et la fécondité*. CEPREMAP .
- Piketty, T., & Valdenaire, M. (2006). L'impact de la taille des classes sur la réussite scolaire dans les écoles, collège et lycées français. estimations à partir du panel primaire 1997 et du panel secondaire 1995. In M. d. nationale., *Les Dossier* (Vol. 173).

- Piquée, C., & Suchaut, B. (2002). Les opinions des acteurs confrontées à l'évaluation externe. Le cas d'une action d'accompagnement scolaire au CP. *Les sciences de l'éducation - Pour l'ère nouvelle*, 3, 103-127.
- Pisani-Ferry, J. (2016). Les enjeux de l'évaluation. *deuxième journée des Rencontres des acteurs publiques*. Paris: Cour des comptes.
- Rauscher, O., Schober, C., & Millner, R. (2012). *Social Impact Measurement und Social Return on Investment (SROI)-Analysis New methods of economic evaluation?* Vienna: Wirtschafts Universität Wien.
- Riegert, A., Maurin, E., & Ly, S. T. (2016). *Elite Tutors and Underprivileged High-School Students*. Paris: Paris School of Economics.
- Romer, P. (1986). Increasing Returns and Long Run Growth. *Journal of Political Economy*.
- Rychetnik, L., Fromme, r. M., P, H., & shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*, 56, 119-27.
- Saez, E. (2000). *Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses*. National Bureau of Economic Research, Inc.
- Saez, E. (2004, March). Direct or indirect tax instruments for redistribution: short-run versus long-run. *Journal of Public Economics*, 88(3-4), 503-518.
- Sen, A. (1985). *Commodities and capabilities*. Amsterdam: Elsevier.
- SROI Network. (2015). *SROI and Cost Benefit Analysis: Spot the Difference, or Chalk and Cheese?* London: http://socialvalueuk.org/publications/publications/doc_download/620-sroi-and-cost-benefit-analysis.
- Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasiexperimental designs: Matching Methods for Causal Inference. In J. W. Osborne, *Best Practices in Quantitative Methods*.
- Studer, N. (2011). méthodes d'évaluation des politiques publiques. actes du séminaire.
- Suziedelyte, A., & Zhu, A. (2015). Does early schooling narrow outcome gaps for advantaged and disadvantaged children ? . *Economics of Education review*, 76-88.
- The Sroi Network. (2012). *A guide to Social Return on Investment*.
- Thibault, F. (2014). Ambitions et évaluation du Revenu de solidarité active . *Économie et Statistique*, 467-468.
- U.S. Department of Health and Human Services (1996). *Cost-Effectiveness in Health and Medicine: Report to the U.S.* . 1996. Public Health Service by the Panel on Cost-Effectiveness in Health and Medicine.
- Urquiola, M. (2006). Identifying class size effects in developing countries : evidence from rural Bolivia. . *Review of Economics and Statistics* , 88(1).
- Vaalavuo, M. (2013). The Redistributive Impact of 'Old' and 'New' Social Spending . *Journal of Social Policy* , 42(3), 513-539.
- Vandenbroucke, F., & Vleminckx, K. (2011). Disappointing poverty trends: is the social investment state to blame? *Journal of European Social Policy*, 21(5), 450-471.
- Vining, A., & Weimer, D. L. (2010). An assessment o fimportant issues concerning the application of Benefit-cost Analysis to Social Policy. *Journal of Benefit-Cost Analysis*, 1(1).
- Vliet, O. v., & Wang, C. (2015). Social Investment and Poverty Reduction: A Comparative Analysis across Fifteen European Countries. *Journal of Social Policy* , 44(3), 611-638.
- Wachter, T. M., & Sullivan, D. (2009, 9). Job Displacement and Mortality: An Analysis using Administrative Data. *Quarterly Journal of Economics*, 1265-1306.
- Wachter, T. M., Schmieder, J., & Bender, S. (2009). *The Long-Term Impact of Job Displacement in Germany During the 1982 Recession on Earnings, Income, and Employment*. Columbia University, Department of Economics.
- Washington State Institute for Public Policy. (2016, June). *Benefit-Cost technical Documentation*. Olympia (WA): Autor.

- Weimer, D. L., & Vining, A. R. (2009). *Investing in the disadvantaged. Assessing the benefits and costs of social policies*. Washington DC: Georgetown University Press.
- Wilson, S. J., Tanner-Smith, E. E., Lipsey, M. W., Steinka-Fry, K., & Morrison, J. (2011). Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *Campbell Systematic Reviews*, 8.
- Wooldridge, J. M. (2009). Omitted Variable Bias: The Simple Case . In J. M. Wooldridge, *Introductory Econometrics: A Modern Approach* (pp. 89-93). Mason, OH: Cengage Learning.

RETROUVEZ
LES DERNIÈRES ACTUALITÉS
DE FRANCE STRATÉGIE SUR :



www.strategie.gouv.fr



[francestrategie](https://www.facebook.com/francestrategie)



[@Strategie_Gouv](https://twitter.com/Strategie_Gouv)



FRANCE STRATÉGIE

France Stratégie, laboratoire d'idées public, a pour mission d'éclairer les choix collectifs. Son action repose sur quatre métiers : évaluer les politiques publiques ; anticiper les mutations à venir dans les domaines économiques, sociétaux ou techniques ; débattre avec les experts et les acteurs français et internationaux ; proposer des recommandations aux pouvoirs publics nationaux, territoriaux et européens. Pour enrichir ses analyses et affiner ses propositions France Stratégie s'attache à dialoguer avec les partenaires sociaux et la société civile. France Stratégie mise sur la transversalité en animant un réseau de huit organismes aux compétences spécialisées.
