



HAL
open science

L'apport des méthodes mixtes à l'évaluation. Une expérimentation randomisée dans le champ des politiques éducatives.

Clément Pin, Carlo Barone

► To cite this version:

Clément Pin, Carlo Barone. L'apport des méthodes mixtes à l'évaluation. Une expérimentation randomisée dans le champ des politiques éducatives.. Revue Française de Science Politique, 2021, Vol. 71 (3), pp.391-412. 10.3917/rfsp.713.0391 . hal-03412051

HAL Id: hal-03412051

<https://hal.science/hal-03412051>

Submitted on 14 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'APPORT DES MÉTHODES MIXTES À L'ÉVALUATION UNE EXPÉRIMENTATION RANDOMISÉE DANS LE CHAMP DES POLITIQUES ÉDUCATIVES

Clément Pin, Carlo Barone

Revue française de science politique – vol. 71, n° 3, 2021

Si l'appel à dépasser le clivage entre méthodes « quantitatives » et méthodes « qualitatives » n'est pas nouveau dans le champ de la science politique, le développement de travaux pratiquant ou revendiquant leur articulation au sein d'un même projet de recherche s'est affirmé ces dernières années¹. Cet article présente un projet de ce type et le mobilise comme un cas permettant de nourrir une réflexion méthodologique centrée sur le recours aux méthodes mixtes dans l'étude de l'action publique. Plus particulièrement, dans le cadre de ce dossier, il s'agit d'illustrer deux stratégies de méthodes mixtes figurant dans la typologie de l'article introductif : celles de la « systématisation » et de l'« enrichissement ».

Du côté de l'action publique, les formes de légitimation des interventions gouvernementales se sont renouvelées ces trente dernières années en intégrant davantage les exigences du pilotage par la performance par la production d'indicateurs chiffrés et la réalisation d'études d'impact². Du côté des sciences sociales, un effort accru a été mené pour produire des inférences causales et mettre au jour les mécanismes causaux à l'œuvre dans les phénomènes observés, aussi bien par des chercheurs de tradition quantitative souhaitant rapprocher les sciences sociales des canons des sciences expérimentales que par des chercheurs de tradition qualitative, soucieux de réaffirmer la dimension théorique de leurs recherches et de mieux contrôler les opérations de montée en généralité. À la croisée de ces évolutions, la pratique de l'évaluation par expérimentation aléatoire contrôlée (*randomized controled trial*) a gagné en popularité³, et avec elle le développement de méthodes expérimentales ou quasi

¹ Voir l'article de Thomas Aguilera et Tom Chevalier dans ce dossier.

² Christopher Hood, « Public Management by Numbers as a Performance-Enhancing Drug : Two Hypotheses », *Public Administration Review*, 72 (1), 2012, p. S85-S92 ; Philippe Bezes, Ève Chiapello, Pierre Desmarez, « Introduction : la tension savoirs-pouvoirs à l'épreuve du gouvernement par les indicateurs de performance », *Sociologie du travail*, 58 (4), 2016, p. 347-369.

³ Cette affirmation vaut essentiellement pour la science politique et pour l'économie, beaucoup moins pour la sociologie. Voir, sur ce point, Michelle Jackson, David R. Cox, « The Principles of Experimental Design and Their Application in Sociology », *Annual Review of Sociology*, 39, 2013, p. 27-49 ; Delia Baldassarri, Maria

expérimentales appliquées aux politiques publiques, comme l'atteste la reconnaissance des travaux d'Esther Duflo, prix Nobel d'économie en 2019.

Ce nouvel engouement pour l'évaluation à dominante quantitative ne manque pas de susciter des controverses dans des disciplines telles que la science politique ou la sociologie, davantage marquées historiquement par le pluralisme méthodologique que ne l'est l'économie. Une part importante des discussions se concentre sur la rhétorique de l'*evidence-based policy*, c'est-à-dire sur les discours promouvant l'usage des connaissances scientifiques pour fonder les politiques publiques sur des données probantes. Des politistes tels que Paul Cairney⁴ ou Brian Head⁵ soulignent que ces discours confinent à la naïveté lorsqu'ils présupposent, en ignorant les acquis des travaux à dominante qualitative sur les processus décisionnels et la mise en œuvre, qu'il suffirait que des connaissances aient une valeur scientifique pour qu'elles s'imposent comme référence dans l'action publique. Pour autant, ces mêmes auteurs cherchent à éviter l'écueil inverse d'une conception « cynique » de l'action publique selon laquelle les luttes institutionnelles, les conflits d'idées et d'intérêts ainsi que les normes et routines qui la caractérisent la rendraient par nature imperméable à toute utilisation des résultats de la recherche. Ils considèrent plutôt que les outils théoriques propres à la science politique peuvent contribuer à une pratique « réaliste » de l'*evidence-based policy*, inscrivant de la sorte leur réflexion dans les pas du sociologue Ray Pawson⁶, qui a lui-même théorisé une démarche « réaliste » d'évaluation des politiques publiques. Il s'agit d'une approche d'évaluation aujourd'hui bien établie scientifiquement et mobilisée par de nombreuses institutions gouvernementales. Elle suggère de substituer à la question habituelle de l'efficacité d'une politique (fonctionne-t-elle, à savoir produit-elle les effets escomptés et prévus ?) un questionnement sur ce qu'elle produit (dans quelles conditions ces effets sont-ils produits ? comment ? pour quels publics ? dans quels contextes ?).

Le présent article propose de se référer au cadre théorique de l'évaluation réaliste et au type de stratégie de recherche qui lui correspond en vue d'explicitier l'apport potentiel des

Abascal, « Field Experiments across the Social Sciences », *Annual Review of Sociology*, 43, 2017, p. 41-73.

⁴ Paul Cairney, *The Politics of Evidence-Based Policy Making*, Basingstoke, Palgrave Macmillan, 2016.

⁵ Brian W. Head, « Reconsidering Evidence-Based Policy : Key Issues and Challenges », *Policy and Society*, 29 (2), 2010, p. 77-94 ; Brian W. Head, « Toward more “Evidence-Informed” Policy Making ? », *Public Administration Review*, 76 (3), 2016, p. 472-484.

⁶ Ray Pawson, Nicholas Tilley, *Realistic Evaluation*, Londres, Sage, 1997 ; Ray Pawson, *Evidence-Based Policy. A Realist Perspective*, Londres, Sage, 2006.

méthodes mixtes pour la science politique. À la question de savoir en quoi consiste le fait de produire une expérimentation randomisée en mobilisant des méthodes qualitatives, on répondra que cela s'inscrit dans une démarche de recherche évaluative visant à combiner la recherche d'inférence causale avec la prise en compte des variables contextuelles. Quant à la question de savoir comment peuvent être concrètement articulées méthodes quantitatives et méthodes qualitatives, la recherche présentée dans l'article permettra d'illustrer que cela peut se faire d'au moins deux manières : dans une stratégie de « systématisation », où les méthodes qualitatives sont mobilisées en amont d'une expérimentation randomisée (quantitative) pour aider à sa conception – on parlera alors de « qualitatif instrumentalisé » ; et dans une stratégie d'« enrichissement », où les méthodes qualitatives sont utilisées ultérieurement à l'expérimentation randomisée, pour répondre à des besoins nouveaux de théorisation issus de l'expérimentation – on parlera cette fois de « qualitatif autonomisé ».

Les deux stratégies de recherche seront illustrées à partir du projet nommé « *Language Skills*⁷ » dont l'objet était de concevoir, mettre en œuvre et évaluer un programme d'action de lecture parentale visant à favoriser le développement des compétences langagières d'enfants de milieux défavorisés et, à terme, leur réussite scolaire. Le concept de « compétences langagières » sert à désigner des phénomènes tels que l'ampleur du vocabulaire, la conscience phonologique et la familiarité avec les conventions de l'écrit. On sait que l'acquisition de ces compétences se joue avant l'école primaire et qu'elle est socialement conditionnée⁸, car le développement langagier dépend fortement des interactions entre parents et enfants : plus les parents sont instruits, plus ils consacrent de temps à des activités de stimulation de leurs enfants et plus ils veillent à intégrer dans leur quotidien des activités éducatives intentionnelles ainsi que des activités d'apprentissage informel tels que la lecture parentale (parents lisant à leurs enfants). On sait par ailleurs que les compétences langagières conditionnent de nombreux autres apprentissages et qu'elles sont un prédicteur important des inégalités de réussite scolaire⁹. La

⁷ Le nom exact du projet est « Favoriser les compétences langagières des enfants des milieux défavorisés : une évaluation expérimentale d'un dispositif d'accompagnement à la lecture parentale ». Il a été conçu et mis en œuvre entre septembre 2016 et juin 2018 sous la direction d'une équipe de sociologues et d'économistes de l'éducation composée de Carlo Barone, Denis Fougère et Agnès van Zanten. Cette recherche a bénéficié du soutien apporté par l'ANR et l'État au titre du programme d'Investissements d'avenir dans le cadre du labex LIEPP (ANR11LABX0091, ANR 11 IDEX000502) dans le cadre d'un appel d'offre de projets internes.

⁸ Lay See Yeo, Winston W. Ong, Charis M. Ng, « The Home Literacy Environment and Preschool Children's Reading Skills and Interest », *Early Education and Development*, 25 (6), 2014, p. 791-814.

⁹ National Institute of Child Health and Human Development, « Developing Early Literacy. Report of the

lecture parentale peut ainsi être considérée comme un médiateur des inégalités sociales dans l'acquisition des compétences langagières ainsi que dans la réussite scolaire.

Le projet « *Language Skills* » consistait à réaliser une évaluation *ex ante*, dont les grandes étapes ont été les suivantes. De septembre à décembre 2016 une enquête qualitative a été réalisée auprès de parents et de directrices d'écoles maternelles pour concevoir la forme finale et le contenu concret du programme. Sur la même période, un tirage aléatoire de 22 écoles maternelles en Réseau d'éducation prioritaire (REP et REP+) dans les XVIII^e, XIX^e et XX^e arrondissements de Paris a été effectué, et, en janvier 2017, un pré-test de vocabulaire a été administré à 900 élèves de moyenne section de ces écoles. À la suite du pré-test, dans chaque école, la moitié des classes a été assignée au groupe de traitement et l'autre moitié au groupe de contrôle. De février à mai 2017, le groupe de traitement a bénéficié du programme, qui contenait quatre éléments : le prêt de deux livres par semaine à chaque enfant, la fourniture aux parents de brochures présentant des informations et des conseils sur la pratique de lecture parentale, la réalisation d'un appel téléphonique aux parents au moment du lancement du programme afin de les impliquer dans le dispositif et leur communiquer quelques messages clés figurant dans les brochures, enfin l'envoi de plusieurs SMS aux parents, en fin de programme, reprenant de manière synthétique des informations et conseils contenus dans les brochures. Les post-tests de vocabulaire ont été réalisés en juin 2017 et en novembre 2017 (*follow-up*). Une deuxième vague d'expérimentation sur 24 nouvelles écoles a été menée l'année scolaire suivante (2017-2018) suivant la même méthodologie afin de répliquer les résultats et d'augmenter la taille globale de l'échantillon de l'étude.

L'analyse des données de l'expérimentation montre que le programme a eu des résultats significatifs sur la fréquence de la lecture partagée ainsi que sur le développement langagier des enfants, en particulier sur les enfants de milieux défavorisés¹⁰. Toutefois, comme nous le soulignerons, l'expérimentation n'ayant pas eu les mêmes effets auprès des différentes sous-catégories de public ciblé, la compréhension de ses effets gagne à être enrichie d'une analyse

National Early Literacy Panel », Washington, US Government Printing Office, 2010.

¹⁰ Pour une présentation complète des résultats, nous renvoyons à Carlo Barone, Denis Fougère, Clément Pin, « Social origins, shared book reading and language skills in early childhood: evidence from an information experiment », *European Sociological Review*, 37 (1), 2020, p. 18-31. Pour une présentation synthétique, voir Carlo Barone, Denis Fougère D., Clément Pin, « La lecture partagée : un levier pour réduire les inégalités scolaires ? L'évaluation par expérimentation aléatoire d'un dispositif dans des écoles maternelles », *LIEPP Policy Brief*, n° 44, 2019.

qualitative des modalités de réception variées¹¹ du programme selon les familles. À cette fin, nous avons réutilisé les entretiens que nous avons réalisés avec des parents dans le cadre de l'enquête qualitative préalable, dont nous nous étions déjà servis pour concevoir le programme d'action, afin, cette fois, de produire une typologie du comportement des parents à l'égard de notre programme permettant de mieux comprendre la diversité de ses effets. Le projet « *Language Skills* » a ainsi permis d'opérer un qualitatif « à double détente », en amont puis en aval de l'expérimentation. Plutôt que de soutenir la thèse d'une contradiction de principe entre la démarche d'évaluation réaliste et la pratique de l'expérimentation randomisée, le présent article souligne à l'inverse la possibilité et l'intérêt de les combiner.

Ce texte commencera par dresser un état de l'évaluation scientifique dans le champ des politiques éducatives et en particulier des apports et limites du recours à l'expérimentation randomisée, afin de mettre en perspective la stratégie de recherche globale du projet « *Language Skills* ». La deuxième partie détaillera les enjeux et modalités de l'articulation des méthodes dans le cadre d'une stratégie de systématisation : il s'agira de souligner les caractéristiques méthodologiques d'une expérimentation de terrain (*field experiment*) et la manière spécifique dont cette pratique traite les enjeux de validité interne des résultats d'une recherche. La troisième et dernière partie portera sur la conduite d'une stratégie d'enrichissement en illustrant celle-ci par un travail de théorisation, en l'occurrence la production d'idéaux-types, sur lesquels peuvent déboucher les résultats d'une évaluation par expérimentation randomisée lorsque celle-ci s'inscrit dans une démarche d'évaluation réaliste et qu'elle mobilise les méthodes mixtes.

À la recherche d'un effet et de sa cause : la stratégie de recherche d'une évaluation par méthodes mixtes

Cette partie vise à présenter la stratégie de recherche globale du projet « *Language Skills* », afin de rendre compte de la perspective théorique de l'évaluation réaliste dans laquelle ont été combinées méthode quantitative et méthode qualitative.

La lecture parentale au prisme des limites de l'*evidence-based policy* dans les politiques éducatives

¹¹ Anne Revillard, « Saisir les conséquences d'une politique à partir de ses ressortissants : la réception de l'action publique », *Revue française de science politique*, 68 (3), juin 2018, p. 469-491.

Bien que l'évaluation figurait dans le projet initial des *policy sciences* incarnées par Harold Lasswell, un clivage épistémologique durable s'est par la suite opéré entre la science politique et l'évaluation. Marqués par le paradigme de la rationalité limitée des travaux d'Herbert Simon¹² ainsi que par les analyses de la mise en œuvre inaugurées par Jeffrey L. Pressman et Aaron Wildawsky¹³, les politistes ont eu tendance à considérer qu'il était en soi problématique d'attribuer des effets ou des impacts aux politiques publiques en tant que telles puisqu'un tel raisonnement s'arrimait à un postulat positiviste. L'évaluation, au sens de l'« évaluation de programme », s'est quant à elle constituée comme un champ intellectuel largement autonome de la science politique¹⁴, en mobilisant d'autres disciplines telles que la psychologie, autour de Donald Campbell¹⁵, ou la santé publique, avec Archibald Cochrane¹⁶.

Malgré la promotion du pluralisme méthodologique puis d'une valorisation grandissante des méthodes mixtes au sein de la science politique, on constate que l'étude des *outcomes*, à la différence des *outputs*, y reste encore largement sous-investie¹⁷. Dans le champ de l'action publique, en revanche, la rhétorique de l'*evidence-based policy*, qui s'est développée dans les années 1990 aux États-Unis et au Royaume-Uni puis qui s'est diffusée dans les années 2000 aux pays de l'Organisation de coopération et de développement économiques (OCDE) et de l'Union européenne, a eu pour effet d'accroître l'intérêt pour les démarches d'évaluation à dominante quantitative telles que l'expérimentation randomisée, les revues systématiques de la littérature et les méta-analyses de données¹⁸. Ce mouvement est porté par des consortiums internationaux, tels que la Cochrane Collaboration dans le secteur de la santé ou la Campbell

¹² Herbert A. Simon, *Administrative Behavior. A Study of Decision-Making Processes in Administrative Organizations*, New York, The Free Press, 1947 ; James G. March, Herbert A. Simon, *Organizations*, New York, John Wiley and Sons, 1958.

¹³ Jeffrey L. Pressman, Aaron Wildawsky, *Implementation*, Berkeley, University of California Press, 1973.

¹⁴ D'abord aux États-Unis, au Canada et au Grande-Bretagne, plus tardivement en France. Voir sur ce point Vincent Spenlehauer, *Des sciences sociales engagées. L'évaluation des politiques publiques*, habilitation à diriger des recherches, université Paris-Est, Créteil, 2011.

¹⁵ Donald T. Campbell, « Reforms as experiments », *American Psychologist*, 24 (4), 1969, p. 409-429.

¹⁶ Archie L. Cochrane, *Effectiveness and Efficiency. Random Reflections on Health Services*, Londres, Nuffield Hospitals Trust, 1972.

¹⁷ Patrice Duran, *Penser l'action publique*, Paris, LGDJ, 1999, p. 55-65 ; Pierre Lascoumes, Patrick Le Galès, *Sociologie de l'action publique*, Paris, Armand Colin, 2012 [2007], p. 44.

¹⁸ Romuald Normand, « La circulation internationale des techniques de la preuve dans les programmes d'intervention et de gestion publique en éducation », *Revue française d'administration publique*, 161, 2017, p. 19-30.

Collaboration dans le champ des politiques sociales et éducatives, qui soutiennent la production de « données probantes » permettant de mesurer l'impact des programmes d'action. Le fait que l'approche ethnographique et les études de cas, que nombre de politistes et sociologues continuent de privilégier, ne figurent pas parmi les méthodes d'évaluation les plus valorisées par ces consortiums ne joue pas en faveur d'une contribution massive de la science politique ou de la sociologie à l'évaluation.

Pour autant, la rhétorique de l'*evidence-based policy* contribue à mettre à l'agenda politique des sujets qui intéressent la science politique ou la sociologie et leur pluralisme méthodologique. Il en va ainsi de l'investissement social dans la petite enfance qui est promue à l'échelle tant internationale que nationale¹⁹. Certes, les promoteurs d'une telle orientation des politiques sociales et éducatives mobilisent en partie des données produites à l'aide de démarches expérimentales ou quasi expérimentales. Mais les arguments mis en avant sont de natures variées : quand certains, plutôt économistes, recourent dans une perspective essentiellement budgétaire et financière à des études d'impact pour étayer un raisonnement en termes d'optimisation des dépenses publiques²⁰, d'autres, en référence à des objectifs de justice sociale, fondent avant tout leur propos sur le constat quantitativement documenté de la dimension cumulative tant des processus de développement cognitif (comme le font certains psychologues ou neuroscientifiques) que de la production des inégalités sociales (comme le font plutôt sociologues ou politistes)²¹.

La production de tels arguments n'est pas sans lien avec l'orientation récente des politiques éducatives d'un pays tel que la France, où les enquêtes réalisées par le Programme international pour le suivi des acquis des élèves (PISA) soulignent de manière répétée non seulement le niveau moyen mais aussi le caractère socialement très inégalitaire des compétences acquises par les élèves français²². Ces arguments ne sont pas non plus sans lien avec le consensus politique et scientifique dont jouissent à l'échelle internationale les programmes promouvant auprès des parents la pratique de la lecture avec leurs enfants dès le

¹⁹ R. Normand, « La circulation internationale... », art. cité ; Cyprien Avenel *et al.*, *L'investissement social. Quelle stratégie pour la France ?*, Paris, La Documentation française, 2017.

²⁰ James Heckman, *Giving Kids a Fair Chance (A Strategy That Works)*, Boston, MIT Press, 2013.

²¹ Gøsta Esping-Andersen, Bruno Palier, *Trois leçons sur l'État-providence*, Paris, Seuil, 2008.

²² Denis Meuret, Sophie Morlaix, « L'influence de l'origine sociale sur les performances scolaires : par où passe-t-elle ? », *Revue française de sociologie*, 47 (1), 2006, p. 49-79 ; Marielle Le Mener, Denis Meuret, Sophie Morlaix, « L'accroissement de l'effet de l'origine sociale sur la performance scolaire : par où est-il passé ? », *Revue française de sociologie*, 58 (2), 2017, p. 207-231.

plus jeune âge²³. Ces programmes dits de lecture parentale, dont les plus connus d'entre eux sont les programmes « *Book start* » et « *Reach out and read* » aux États-Unis, se sont multipliés ces dernières années dans l'objectif affiché de réduire les inégalités sociales en matière d'éducation²⁴. Pourtant, à la suite d'études antérieures interrogeant l'efficacité de la lecture parentale²⁵, une récente méta-analyse²⁶ a permis de souligner que les essais randomisés contrôlés portant sur ces programmes de lecture parentale étaient rares (une trentaine) et que, contrairement à l'engouement qu'ils suscitent, les effets de ces programmes sont rarement positifs, exceptés pour un type particulier d'entre eux, à savoir ceux qui privilégient la technique de la « lecture dialogique », une approche qui insiste davantage sur les interactions verbales entre parents et enfants – à la suite de la lecture de l'histoire afin de mobiliser activement le vocabulaire des enfants – que les programmes eux-mêmes. Toutefois, ces programmes de lecture dialogique bénéficient davantage aux enfants de familles favorisées et ont donc tendance à renforcer l'effet des inégalités sociales sur le développement des compétences langagières des enfants. Les évaluations d'impact menées pendant les deux dernières décennies ont ainsi questionné le consensus sur l'efficacité des programmes de lecture parentale et sur leur potentiel égalisateur.

Ces résultats conduisent à interroger non plus seulement l'efficacité des politiques éducatives mais aussi la qualité et la fiabilité des expérimentations randomisées réalisées²⁷. En effet, le faible taux de résultats positifs des expérimentations randomisées peut s'expliquer de deux manières principales. Premièrement, il est possible que les expérimentations menées s'avèrent inefficaces parce qu'elles ont été mal conçues ou mal mises en œuvre, en ce sens que les équipes de recherche qui les portent ne consacrent pas suffisamment d'attention aux conditions concrètes de leur transposition locale. Deuxièmement, il est possible que les

²³ David K. Dickinson *et al.*, « How Reading Books Fosters Language Development Around the World », *Child Development Research*, 2012.

²⁴ Luisa Araújo, Patricia Costa, « Home Book Reading and Reading Achievement in EU Countries : The Progress in International Reading Literacy Study 2011 (PIRLS) », *Educational Research and Evaluation*, 21 (5-6), 2015, p. 422-438.

²⁵ Holis S. Scarborough, Wanda Dobrich, « On the Efficacy of Reading to Preschoolers », *Developmental Review*, 14 (3), 1994, p. 245-302 ; Sharon Goldfeld *et al.*, « Outcomes of a Universal Shared Reading Intervention by 2 Years of Age : the Let's Read Trial », *Pediatrics*, 127 (3), 2011, p. 445-453.

²⁶ Carlo Barone *et al.*, « Shared Book Reading Interventions and Children's Skills : A Meta-Analysis of Randomised Controlled Trials », Sciences Po LIEPP, *Working Paper*, 98, 2019.

²⁷ Hugues Lortie-Forgues, Matthew Inglis, « Rigorous Large-Scale Educational RCTs are often Uninformative : Should we be Concerned ? », *Educational Researcher*, 48 (3), 2019, p. 158-166.

programmes soient efficaces mais que le protocole d'évaluation adopté ne permette pas de détecter leurs effets (propres) de manière fiable, par exemple parce que ces études reposent sur des échantillons trop petits et spécifiques, et présentent donc des limites en termes de validité externe. Le présent article, en portant sur le projet « *Language Skills* », vise à examiner les deux facteurs mentionnés. On commencera pour ce faire par souligner qu'il s'agit d'une recherche qui, tout en recourant à une expérimentation randomisée, souscrit aux principes d'une démarche d'évaluation réaliste.

Une démarche d'évaluation réaliste intégrant un protocole expérimental

Au regard de la littérature sur les programmes de lecture parentale et plus particulièrement sur leur évaluation par expérimentation randomisée, la stratégie globale du projet « *Language Skills* » visait un double objectif scientifique : prouver qu'un programme de lecture parentale pouvait être efficace et montrer que cette efficacité pouvait être conciliable avec des objectifs de justice sociale généralement affichés par les gouvernements. À ce titre, le projet a suivi une démarche d'évaluation réaliste, celle-ci présentant la caractéristique de combiner une double exigence : celle d'explicitier la théorie du changement sous-jacente d'un programme d'action et celle d'être attentif à la diversité des effets qu'un même programme est susceptible de produire selon les publics bénéficiaires et plus largement selon les contextes de réception²⁸.

Concernant l'explicitation de la théorie causale d'un programme, l'évaluation réaliste s'inspire de « l'évaluation guidée par la théorie » conceptualisée par Huey T. Chen²⁹ et Carol H. Weiss³⁰. Reconstituer cette théorie consiste à décomposer analytiquement les hypothèses sur lesquelles se fonde un programme et à les inscrire dans un enchaînement causal en quatre étapes³¹ : 1. l'étape des « réalisations », c'est-à-dire des activités mises en œuvre ; 2. celle des résultats au sens d'accès effectif des publics ciblés au programme (*outputs*) ; 3. celle des impacts intermédiaires, autrement dit l'impact qu'est supposé avoir le programme sur les

²⁸ Cette double exigence est bien soulignée dans Émilie Robert, Valéry Ridde, « L'approche réaliste pour l'évaluation de programmes et la revue systématique : de la théorie à la pratique », *Mesure et évaluation en éducation*, 36 (3), 2013, p. 79-108.

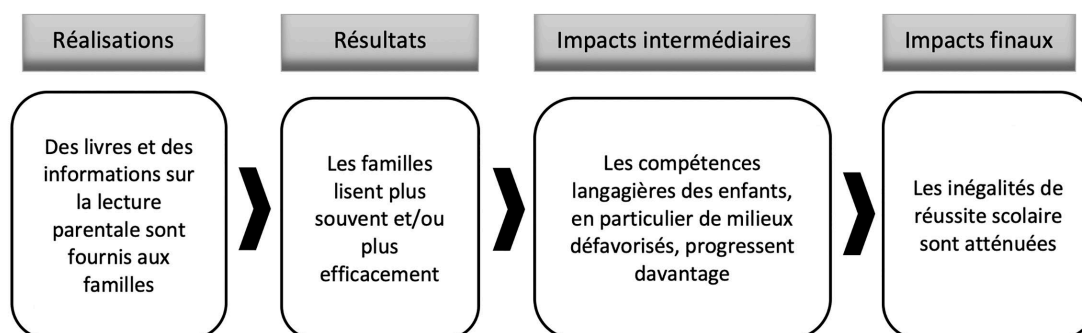
²⁹ Huey T. Chen, *Theory-Driven Evaluations*, Beverly Hills, Sage, 1990.

³⁰ Carol H. Weiss, « How Can Theory-Based Evaluation Make Greater Headway ? », *Evaluation Review*, 21 (4), 1997, p. 501-524.

³¹ Sue C. Funnell, Patricia J. Rogers, *Purposeful Program Theory. Effective use of Theories of Change and Logic Models*, Chichester, Wiley, 2011.

bénéficiaires (*outcomes 1*) ; 4. les « impacts finaux », soit la finalité escomptée de l'intervention (*outcomes 2*). Dans le cas du projet « *Language Skills* » (voir figure 1), la réalisation supposée du programme était la fourniture aux familles de livres et d'informations sur la pratique de la lecture parentale. Le résultat attendu était que les familles bénéficiaires du programme utilisent les livres et les informations qui leur ont été transmises pour lire davantage et/ou plus efficacement qu'elles ne le faisaient auparavant. L'impact intermédiaire supposé était que les compétences langagières des enfants s'accroissent, en particulier celles des enfants de milieux défavorisés. L'impact final supposé, enfin, était que, à plus long terme, grâce à un meilleur développement précoce des compétences langagières des enfants de milieux défavorisés, les inégalités de réussite scolaire tendaient à décroître. Il convient de souligner que si nous avons effectué une deuxième vague de tests une année plus tard, l'impact final n'a pas pu être étudié à long terme puisque l'acceptabilité par les écoles de la randomisation effectuée en leur sein (entre classes) a nécessité que nous proposons un traitement différé (*delayed treatment*) pour les classes témoins. Cette absence de mesure des impacts finaux et la focalisation sur le court terme constituent une limite méthodologique souvent reprochée aux expérimentations randomisées. Par exemple, dans notre cas, il se pourrait que les enfants du groupe traité oublient les mots appris ou que ceux du groupe contrôle les aient appris par la suite. Toutefois, à six mois de distance, avant le *delayed treatment*, nous avons observé que les effets du programme persistent et qu'ils sont même renforcés.

Figure 1. Logigramme représentant la théorie du programme dans le projet « *Language Skills* »



Source : adapté de Devaux-Spatarakis³² ; conception des auteurs.

³² Agathe Devaux-Spatarakis, « L'évaluation 'basée sur la théorie', entre rigueur scientifique et contexte politique », *Politiques et management public* [En ligne], 31 (1), 2014.

Quant à l'intention de contextualiser les effets générés par le programme, elle conduit à étudier les composantes réelles et non plus hypothétiques d'un programme. Dans le projet « *Language Skills* », le mécanisme que l'on cherchait à activer était un changement de comportement des parents en matière de lecture parentale, en levant une part des barrières informatives qui s'exercent sur eux à propos de cette activité, en termes d'utilité et de pratique (voir encadré 1). Le concept de mécanisme occupe une place centrale dans l'opérationnalisation de la démarche d'évaluation réaliste, en désignant non pas le programme, mais un processus activé par la mise en œuvre d'un programme dans un contexte donné³³. L'utilité du concept tient au fait qu'il ne suffit pas de constater un changement de situation après le lancement d'un programme pour en déduire que le changement constaté est un impact du programme. Il faut pour ce faire documenter le phénomène d'activation (ou de non activation) du mécanisme par le programme, dans des contextes différents, pour pouvoir suivre ensuite les effets qu'il génère (ou non) et qui se répercutent sur les impacts intermédiaires et finaux. Une évaluation réaliste ne se contente donc pas de reconstituer la théorie sous-jacente d'un programme : elle accorde également une grande attention à l'étude de son contexte d'application, qu'il s'agisse de son environnement institutionnel, organisationnel, socio-économique ou culturel.

Encadré 1. Le système d'hypothèses au cœur du projet « *Language Skills* »

L'hypothèse générale est que les informations concernant les bénéfices de la lecture partagée sont inégalement réparties entre les groupes sociaux : les parents qui ne disposent pas d'un diplôme du supérieur, mais au mieux d'un diplôme du secondaire, ont moins accès aux sources expertes et aux réseaux informels où cette information circule le plus souvent. Par conséquent, nous émettons l'hypothèse que le fait de leur fournir cette information augmentera leur fréquence de lecture partagée (H1a). Nous supposons que l'information aura moins d'impact sur la fréquence de lecture partagée des parents très scolarisés, qui sont moins exposés aux obstacles à l'information (H1b). De plus, nous faisons l'hypothèse que les compétences langagières des enfants issus de familles moins diplômées au sein du groupe de parents traités s'amélioreront grâce à l'augmentation de la fréquence de la lecture partagée (H2a). Enfin, nous faisons l'hypothèse que les compétences langagières des enfants traités issus de familles les plus diplômées s'amélioreront également (H2b), et ce parce que les styles de lecture interactifs pratiqués par ces parents sont davantage bénéfiques pour les compétences langagières.

³³ Le concept de mécanisme s'inscrit plus largement dans celui de « configuration Contexte-Mécanisme-Effet ». Voir sur ce point Ray Pawson, Nicholas Tilley, « In with the New : Introducing Scientific Realism », dans R. Pawson, N. Tilley, *Realistic Evaluation...*, *op. cit.*, p. 55-82. Sur l'enjeu et l'équivocité du concept de mécanisme dans l'analyse des politiques publiques, voir plus largement James Mahoney, « Beyond Correlational Analysis : Recent Innovations in Theory and Method », *Sociological Forum*, 16 (3), 2001, p. 575-593.

Afin de documenter cette dimension contextuelle dès la phase de conception du programme, l'équipe a réalisé une enquête qualitative par entretien afin de récolter des données utiles pour concevoir la forme définitive et le contenu concret du programme. L'enquête a permis de réaliser 24 entretiens, dont 21 avec des parents et 3 avec des directrices d'école. Ces entretiens ont été effectués avant le tirage au sort des écoles et n'ont pas été menés (ou alors seulement par hasard) avec des interviewés par la suite impliqués dans le programme. La population et le terrain de l'enquête ont toutefois été définis en référence à l'objectif du programme de cibler les élèves de milieux défavorisés : le périmètre retenu a été les XVIII^e, XIX^e et XX^e arrondissements de Paris ainsi que les communes limitrophes (Saint-Ouen et Saint-Denis). Ensuite, au cours de l'expérimentation, l'équipe a assuré le monitoring de sa mise en œuvre dans les écoles, en appréhendant cette activité comme une situation d'observation participante. Ce travail d'observation a été réalisé en deux étapes, l'une en janvier 2017, l'autre entre février et mai 2017. La première visait à utiliser la première visite aux écoles tirées au sort pour réaliser de brefs entretiens avec leur direction ainsi qu'avec les enseignants des classes concernées (classes des groupes de traitement et de contrôle). Pour la deuxième étape, il s'agissait de profiter des visites ultérieures effectuées périodiquement dans les écoles pour recueillir de manière informelle des informations sur le déroulement du programme, les difficultés éventuellement rencontrées, et l'intérêt des élèves et des parents.

Si les principes généraux du programme, quant à sa forme et à son contenu, avaient été définis avant même la réalisation de l'enquête qualitative préalable (choix de passer par les écoles, choix d'utiliser des brochures, des appels téléphoniques et des SMS pour échanger avec les parents et leur transmettre des informations), beaucoup restait à faire pour lui donner sa forme concrète et l'ajuster aux conditions réelles de sa mise en œuvre. Les entretiens avec les parents, de type semi-directif, ont été conduits sur la base d'un guide d'entretien portant sur l'environnement familial de l'enfant, l'organisation du temps extra-scolaire et la pratique de la lecture parentale. Ces entretiens ont été utilisés pour identifier des conseils pratiques que nous avons présentés dans les brochures puis dans les SMS. Les entretiens avec les directions d'écoles visaient, quant à eux, à recueillir des données relatives aux types de projets qu'elles menaient pour promouvoir la lecture, et leurs point de vue et recommandations concernant les modalités de mise en œuvre de notre programme. Cet usage des méthodes qualitatives pour concevoir une démarche expérimentale fournit une bonne illustration de la stratégie de « systématisation » qu'il s'agit maintenant de présenter.

Stratégie de systématisation : le « qualitatif instrumentalisé » au cœur d'un *field experiment*

Il s'agit dans cette partie de souligner que la stratégie de systématisation adoptée dans le projet « *Language Skills* » ne se réduit pas à l'usage d'une méthodologie seulement quantitative mais qu'elle intègre, en amont, un travail de recherche qualitatif « instrumentalisé » pour identifier les conditions de mise en œuvre d'un protocole expérimental dans des conditions pratiques singulières.

La méthodologie d'une expérimentation de terrain

Une expérimentation de terrain et une expérimentation de laboratoire ont en commun de recourir au procédé de l'assignation aléatoire, qui confère une forte validité interne aux inférences causales produites puisqu'il garantit que les deux groupes (traité/témoin) disposent des mêmes caractéristiques initiales. En revanche, une expérimentation de terrain se distingue d'une expérimentation de laboratoire par le fait d'étudier les comportements des individus en conditions naturelles, dans des contextes non artificiels.

Dans le cas de « *Language Skills* », à la différence de nombreux autres projets³⁴, les avantages recherchés à travers la pratique d'une expérimentation de terrain étaient moins un gain en termes de validité externe qu'un gain en termes de connaissance des conditions de variation des résultats selon les types de population et selon les contextes de l'expérimentation. À ce titre, on peut dire que l'intérêt de recourir à une expérimentation de terrain est de produire des résultats disposant d'une meilleure « validité écologique ». La validité écologique tient au fait que la pratique de l'expérimentation de terrain implique de documenter la variabilité de l'impact d'un même programme selon le profil de ses bénéficiaires, et donc de disposer d'une analyse circonstanciée de son efficacité. Cela fait de l'expérimentation de terrain une méthodologie particulièrement utile pour l'évaluation de l'action publique, en particulier dans le cas de programmes d'action ayant pour objectif, dans une logique redistributive, de réduire les inégalités constatées entre différents groupes sociaux.

Si le propre d'une expérimentation de terrain est d'être réalisée dans des conditions « naturelles », sa contrepartie est qu'elle expose les chercheurs qui s'y prêtent à une contrainte particulière : celle de renoncer à une part non négligeable de leur contrôle direct sur l'expérimentation et plus précisément sur l'administration du traitement (c'est-à-dire du

³⁴ *Ibid.*, p. 47.

programme). À cette contrainte répond dès lors un prérequis méthodologique : celui de disposer d'une connaissance approfondie du contexte de l'expérimentation, à travers des méthodes qualitatives, permettant d'une part de statuer, dans la phase de conception du programme, sur le partage des rôles opérationnels entre les chercheurs et les autres acteurs impliqués (acteurs de la mise en œuvre ou bénéficiaires), et d'autre part de suivre la mise en œuvre du programme en vérifiant que les acteurs impliqués respectent les instructions qui leur sont données. Ce dernier point est essentiel car lui seul permet de garantir que le programme se mette bien en œuvre de manière homogène auprès de l'ensemble des bénéficiaires, sans quoi la validité interne de l'évaluation serait altérée.

Concrètement, dans le projet « *Language Skills* », à partir du moment où le choix a été fait de tester l'impact d'un programme en le mettant en œuvre à travers les écoles et en mobilisant leur personnel, on s'exposait au risque qu'une partie des écoles et de leur personnel n'adopte pas le comportement requis, soit en ne mettant pas en œuvre comme prévu le programme destiné au groupe traité (on parle alors de biais de *non-compliance*), soit en faisant bénéficier du traitement, au moins en partie, le groupe contrôle (on parle alors de biais de *treatment contamination*).

Les méthodes qualitatives au service de la conception du traitement employé dans une expérimentation randomisée

Une enquête qualitative réalisée en amont d'une expérimentation a pour fonction de produire une connaissance des conditions de mise en œuvre d'un traitement efficace permettant tout à la fois d'identifier les principaux risques de neutralisation, d'altération ou de détournement pour l'administration du traitement, et d'identifier des leviers permettant de désamorcer ces risques contre lesquels se prémunir autant que possible. Une telle utilisation des méthodes qualitatives, et en particulier de la réalisation d'une enquête par entretien, a ceci de particulier qu'elle n'appréhende pas la phase de collecte de données dans une perspective purement académique de mise au point conceptuelle d'une théorie (*conceptual focusing*) mais dans une perspective plus pratique de recherche expérimentale et d'apprentissage collectif (*teacher-learner function*)³⁵. Ce type d'approche répond à la préoccupation évoquée ci-dessus

³⁵ Ray Pawson, Nicholas Tilley, « How to Construct Realistic Data : Utilizing Stakeholders' Knowledge », dans R. Pawson, N. Tilley, *Realistic Evaluation...*, *op. cit.*, p. 153-182, ici p. 155 ; Ray Pawson, « Theorizing the Interview », *British Journal of Sociology*, 47 (2), 1996, p. 295-314 ; Ana Manzano, « The Craft of Interviewing in Realist Evaluation », *Evaluation*, 22 (3), 2016, p. 342-360.

d'être particulièrement attentif aux conditions de transposition locale d'une expérimentation randomisée, autrement dit de veiller à sa traduction opérationnelle en misant sur une coordination étroite entre les chercheurs et les praticiens impliqués, et plus largement entre les activités de conception et de développement du programme testé.

Dans le projet « *Language Skills* », l'enquête qualitative préalable visait à satisfaire le prérequis méthodologique d'une expérimentation de terrain qui est de favoriser son efficacité en prenant en compte le point de vue des *stakeholders* impliqués. Les entretiens ont permis de documenter trois principaux types d'enjeu.

Les premiers, d'ordre éthique et/ou politique, étaient ceux de l'acceptabilité de la démarche scientifique par l'institution scolaire, et en premier lieu par ses acteurs décisionnaires, dans notre cas les inspectrices de l'Éducation nationale (au niveau de l'académie et de ses circonscriptions) et les directrices d'école³⁶. Le principe de l'égalité de traitement des élèves propre au service public de l'éducation fait par définition obstacle au principe d'une démarche expérimentale reposant sur la différenciation de traitement entre un groupe de traitement et un groupe de contrôle. Les entretiens ont permis de vérifier la vigilance de nos interlocutrices sur cette question, et surtout de tester la convenance à leurs attentes de la solution que nous proposons. Notre proposition a été de garantir les conditions d'un traitement différé : le matériel mis à disposition de l'école dans le cadre de l'expérimentation devenait sa propriété à la fin du projet, si bien qu'elle pouvait le réutiliser pour en faire bénéficier ultérieurement les élèves du groupe de contrôle. Au cours des entretiens, une autre question était systématiquement soulevée par les directrices d'école, celle de savoir si et comment l'intervention prenait en compte la situation particulière des familles allophones, parlant uniquement une langue autre que le français à domicile. Ces familles étaient très minoritaires dans notre échantillon, puisqu'elles ne représentaient que 3,5 % de la population étudiée. En revanche, on pouvait s'attendre à ce que les familles parlant non seulement le français mais aussi une autre langue soient fortement représentées, et que, dans ce cas, prévoir de communiquer avec les parents dans d'autres langues que le français pourrait être perçu comme un signal du caractère inclusif de notre programme, et que cela serait de nature à favoriser une large participation tant des parents que des équipes pédagogiques. Les entretiens avec les directrices ont permis d'identifier cinq langues étrangères (anglais, arabe, chinois, bambara,

³⁶ Pour une présentation synthétique des enjeux éthiques de la pratique de l'expérimentation en sciences sociales, nous renvoyons à Paul J. Gertler *et al.*, *Impact Evaluation in Practice*, Washington, International Bank for Reconstruction and Development/The World Bank, 2019, p. 231-243.

tamoul) dans lesquelles il était pertinent de traduire le formulaire d'adhésion parentale et de réaliser l'appel téléphonique aux parents prévu dans le programme.

Le deuxième type d'enjeu était d'ordre organisationnel et concernait plus directement la mise en œuvre du programme : l'institution scolaire française, écoles maternelles incluses, repose sur le couplage singulier entre une centralisation nationale du fonctionnement administratif et une importante autonomie des personnels enseignants à l'égard de leurs chefs d'établissement³⁷. L'absence de pilotage hiérarchique direct des chefs d'établissement vis-à-vis des enseignants rendait nécessaire d'instaurer un contact direct avec ceux-ci et de les faire adhérer aux principes et modalités de l'expérimentation, indépendamment et même en raison du caractère aléatoire de l'assignation de leur classe au groupe traité ou contrôle. Sur le conseil des directrices d'école, un soin particulier a été accordé au fait de rendre le programme le moins chronophage possible, en fournissant aux classes des dispositifs « clés en main » comprenant non seulement les livres et les brochures à destination des familles, mais aussi les instructions à destination des enseignants pour le fonctionnement du prêt de livres dans leurs classes et la mise à disposition de sacs permettant la circulation des livres entre les écoles et le domicile des enfants. Cet exemple permet de souligner l'apport précieux du « qualitatif instrumentalisé » pour la conception du programme (*treatment design*) : le caractère « clé en main » du programme a été d'autant plus recherché que les entretiens tant avec les parents qu'avec les directrices d'école avaient permis de nous familiariser avec le contexte de notre intervention et en particulier de documenter l'existence et les limites des dispositifs déjà mis en place dans les écoles pour encourager la pratique de la lecture à la maison.

La conception du discours et des informations à diffuser auprès des parents par l'intermédiaire des brochures, SMS et appels téléphoniques a été un troisième apport important du « qualitatif instrumentalisé ». L'une des principales hypothèses de notre recherche était que les parents de niveau de formation modeste lisent moins avec leurs enfants parce qu'ils n'ont pas accès aux informations concernant les bénéfices de la lecture parentale, mais aussi parce que leurs conditions de vie étant plus précaires, il leur est plus difficile de s'organiser pour disposer d'un moment propice à la lecture avec leurs enfants. Un enjeu décisif des entretiens était ainsi de mieux connaître le quotidien de ce type de familles, la nature des difficultés qu'elles rencontrent pour pratiquer la lecture parentale et les moyens que trouvent certaines

³⁷ Agnès van Zanten, « La construction des politiques d'éducation : de la centralisation à la délégation au local », dans Pepper D. Culpepper, Peter A. Hall, Bruno Palier (dir.), *La France en mutation, 1980-2005*, Paris, Presses de Sciences Po, 2006, p. 229-263, ici p. 247-249.

d'entre elles pour y parvenir. Les témoignages des parents nous ont conduits à souligner que la priorité n'était pas d'effectuer de longues séances de lecture, mais de mettre en place des séances régulières, quitte à ce qu'elles soient courtes, et que l'instauration de ce type de pratique pouvait générer un gain aussi bien pour le développement de leur enfant que pour la gestion du temps domestique en favorisant, par exemple, la mise au lit en fin de journée. Dans le même ordre d'idées, nous avons introduit dans une brochure l'idée que la lecture « parentale » pouvait être déléguée aux grands-frères ou aux grandes-sœurs.

Mesurer la variabilité des effets selon les contextes : un enjeu majeur pour une évaluation par expérimentation randomisée

Un autre point d'attention décisif pour favoriser la qualité d'une expérimentation randomisée réside dans le fait d'utiliser un protocole d'évaluation fiable, permettant de détecter les effets attendus du programme. Cela suppose de disposer d'un instrument de mesure quantitative des effets, mais aussi de documenter des variables contextuelles susceptibles de jouer sur l'activation du mécanisme recherché par le programme. Dans le cas du projet « *Language Skills* », ces deux aspects avaient été pris en compte dès la conception initiale du protocole. Bien qu'ils ne soient pas le produit du « qualitatif instrumentalisé » auquel nous avons eu recours, ils rendent compte de la démarche consistant à intégrer la pratique de l'expérimentation randomisée dans une évaluation de type réaliste et à recourir de ce fait aux méthodes mixtes.

Afin de mesurer l'effet du programme sur l'évolution des compétences langagières des enfants, le choix a été fait d'utiliser un test de vocabulaire, l'*Échelle de vocabulaire en images Peabody* (EVIP)³⁸. Ce test fournit une mesure de l'étendue du vocabulaire réceptif des enfants d'âge préscolaire dont la validité et la fiabilité sont les mieux établies.

La documentation des variables contextuelles a été recueillie en administrant un court questionnaire aux parents. Un premier ensemble de questions devait permettre d'analyser la différenciation des effets du programme selon les caractéristiques des enfants et de leurs familles (âge de l'enfant, nombre de frères et sœurs, langues parlées à la maison, niveau de formation des parents). Un deuxième ensemble de questions visait à permettre de mesurer les « résultats » du programme, c'est-à-dire ses effets en termes de réception. Ces questions ont ainsi été posées deux fois : une fois au moment du pré-test puis en même temps que le post-test. Elles ont permis de recueillir des données déclaratives des parents sur la fréquence de leur

³⁸ *Peabody Picture Vocabulary Test* (PPVT) en anglais.

pratique de lecture parentale et sur l'intérêt, voire le plaisir, qu'eux-mêmes et leurs enfants y trouvent.

L'analyse des données de l'expérimentation a permis de mesurer les effets du programme en termes de réception par les familles et en termes d'impacts intermédiaires sur les compétences langagières (voir tableau 1)³⁹. Sur le premier point, il ressort que le programme a augmenté sensiblement la fréquence de lecture (colonne 1). Alors qu'avant l'intervention 41 % des parents déclaraient lire chaque jour, après l'intervention la probabilité de lire chaque jour des livres est plus élevée de 8 % pour le groupe traité que pour le groupe contrôlé : l'intervention a ainsi permis d'augmenter la fréquence de lecture d'environ un cinquième. On observe toutefois que cet impact varie fortement selon le niveau d'éducation des parents : il est plus fort pour les ménages peu scolarisés, tandis qu'aucun impact statistiquement significatif n'apparaît pour les familles très scolarisées, pour lesquelles la fréquence de lecture était déjà élevée avant le programme. On peut donc conclure que le programme réduit les inégalités sociales d'accès à la lecture partagée. Concernant l'impact sur les compétences langagières, l'analyse confirme (colonne 2) que la mesure relative au vocabulaire est positive et statistiquement significative. L'analyse indique enfin que le programme a eu un impact positif supérieur sur les compétences langagières des enfants des familles les moins diplômées, que sur celles des enfants issus de ménages disposant d'un diplôme du supérieur, pour lesquels l'effet est positif, mais plus faible et non statistiquement significatif. Les résultats de la recherche permettent de valider les hypothèses 1a, 1b et 2a mais pas, en revanche, l'hypothèse 2b (voir encadré 1), rien ne permettant d'exclure l'hypothèse que le programme n'a eu aucun effet sur les enfants dont les parents sont les plus diplômés. Au regard de ces mesures, l'expérimentation a permis de produire des données probantes de l'efficacité du programme conçu et mis en œuvre.

Il faut remarquer que, dans le contexte de cette expérimentation de terrain, il est difficile d'avoir des indications probantes concernant l'implication réelle des parents dans le dispositif proposé (*non-compliance*). Nous ne pouvons pas savoir, par exemple, combien de parents ont effectivement lu les brochures ou les SMS. Les estimations proposées dans le tableau 1

³⁹ L'analyse d'équivalence *ex ante* des deux groupes, avant la mise en œuvre du programme, sur les variables sociodémographiques (genre, âge, niveau d'éducation des parents, origine immigrée), sur des variables relatives à la lecture (fréquence, appréciation des parents et des enfants de cette activité) ainsi que sur les scores de vocabulaire réceptif des enfants, indique que ces deux groupes étaient équivalents avant l'intervention. Les modèles présentés dans le tableau 1 incorporent les scores de vocabulaire réceptif avant l'intervention afin d'augmenter la puissance statistique des modèles et donc la précision des estimations de l'impact du programme.

n'intègrent pas cet aspect et fournissent une estimation de l'impact global du traitement (*intention-to-treat estimators*). Toutefois, nous savons combien de parents du groupe expérimental ont répondu à l'appel téléphonique qui synthétisait les messages principaux de l'intervention : 94 %. Si nous utilisons la réponse à cet appel comme variable instrumentale pour obtenir des estimations de l'impact du programme qui incorporent la *non-compliance* (*average treatment effect on the treated*), les résultats sont presque identiques aux résultats présentés dans le tableau 1. Ce résultat était attendu lorsque la mesure de *compliance* utilisée est si élevée.

Tableau 1. Analyse des données de l'expérimentation randomisée : effets principaux du programme « *Language Skills* » et variations de ces effets en fonction du niveau d'éducation des parents

	Fréquence de lecture (1=chaque jour)	Test de vocabulaire (post-test)
Modèle	Régression binomiale	Régression linéaire
Effet principal du programme	0.08***	0.11**
Effet de l'intervention selon le niveau d'éducation des parents		
Diplôme du supérieur	0.03	0.11
Diplôme du secondaire, ou moins	0.16***	0.13*
Effectif	1,374	1,611
Significativité statistique : ***p<0.01, **p<0.05, *p<0.1		
Lecture : dans ce tableau, pour les régressions binomiales, sont reportés les effets marginaux ; pour les régressions linéaires estimées par moindres carrés ordinaires, sont reportés les coefficients standardisés. Les contrôles sont effectués d'une part pour la même mesure au pré-test, d'autre part pour une variable binaire (1 ^{re} année et 2 ^e année de l'expérimentation). 0.08 signifie que parmi le groupe traité la probabilité de lire chaque jour des livres est plus élevée de 8 % par rapport au groupe contrôle. 0.11 signifie qu'il y a une amélioration de 0,11 % d'un écart-type de la variable dépendante pour la distribution du vocabulaire.		

L'analyse des résultats et des impacts de l'expérimentation atteste de l'importance de prendre en compte les variables contextuelles afin de pouvoir dire non pas seulement si un programme d'action fonctionne, mais ce qui marche dans ce programme (les incitations à lire plus souvent, mais peu l'intention de développer le plaisir pour la lecture), pour qui (surtout pour les enfants dont les parents n'ont pas un niveau de formation supérieur), comment (en passant par les écoles tout en conservant le contrôle des informations transmises aux parents) et à quelles conditions (en fournissant un programme « clé en main » contenant des livres et des informations adaptés aux publics prioritairement ciblés). À ce titre, plutôt que de souligner les différences épistémologiques et méthodologiques entre une démarche expérimentale et une démarche d'évaluation réaliste⁴⁰, notre expérience du projet « *Language Skills* » permet plutôt de souligner la possibilité de les combiner.

⁴⁰ A. Devaux-Spatarakis, art. cité, p. 60.

Stratégie d'enrichissement : le « qualitatif autonomisé » comme ressource de l'évaluation pour la recherche fondamentale

Si un projet de recherche adoptant un protocole expérimental permet d'établir un lien causal entre un programme et ses impacts, ce résultat peut ensuite déboucher sur la conduite d'une stratégie *d'enrichissement* visant à compléter l'analyse par la compréhension des processus causaux ayant abouti aux impacts mesurés. Cette étape complémentaire a été conduite en conférant une valeur autonome aux méthodes qualitatives pour construire des idéaux-types permettant d'imaginer des régularités rendant compte de la diversité des effets qu'un même programme peut avoir.

Des besoins d'enrichissement théorique nés de la mesure des variables contextuelles

Pour la stratégie de méthodes mixtes par enrichissement conduite dans le projet « *Language Skills* » en complément de la stratégie de systématisation, il s'agissait de repartir de la confirmation de l'hypothèse que le programme expérimenté était de nature à produire des effets contrastés selon le type de bénéficiaires. Sur ce point, le volet quantitatif du projet a abouti à deux résultats saillants. Il a en premier lieu permis de valider l'hypothèse que le programme testé avait moins d'impact sur les familles diplômées du supérieur que sur les familles moins diplômées. On peut supposer que ce phénomène tient à un effet de saturation (les familles les plus diplômées pratiquaient déjà fréquemment la lecture parentale). Le deuxième résultat notable est que les effets du programme, bien que positifs, s'avèrent toutefois d'intensité modérée. Ce résultat suggère que l'apport du « qualitatif instrumentalisé » n'a pas été suffisant pour rendre le programme efficace pour un nombre significatif de familles défavorisées. Au terme de la stratégie de systématisation, il nous a donc semblé utile d'approfondir notre connaissance de la variabilité des conditions de réception d'un programme tel que le nôtre. La stratégie d'enrichissement que nous avons conduite a ainsi visé à fournir des explications empiriquement documentées et avec un certain degré de généralité pour rendre intelligible la variabilité des effets observables entre familles défavorisées (*within low-educated*) plutôt qu'entre familles défavorisées et familles favorisées (*between low and high-educated*).

Il convient de préciser que, dans notre cas, le passage de la stratégie de systématisation à celle d'enrichissement n'a pas donné lieu à la collecte de nouvelles données qualitatives, mais uniquement au traitement par des voies différentes des données collectées au cours de l'enquête qualitative préalable. Cette étape de production d'un « qualitatif autonomisé » a procédé d'une

démarche d'évaluation pouvant être rapprochée du courant de l'évaluation qualitative telle que théorisée par Michael Patton⁴¹, qui met « l'accent sur la subjectivité, le point de vue, les pratiques et l'expérience vécue des destinataires » afin de « prendre en considération les significations diverses que différents individus peuvent assigner à un même programme »⁴².

La réutilisation à des fins d'enrichissement de données recueillies en amont de l'expérimentation se justifie par le fait que l'objectif assigné à l'enquête préalable avait en effet été d'interviewer des parents de conditions sociales aussi différenciées que possible. L'hypothèse était qu'au-delà du niveau de formation des parents, qui a été la variable principale prise en compte dans l'analyse des résultats de l'expérimentation, ces mêmes effets seraient susceptibles de varier selon le niveau de précarité des parents et plus largement en fonction des conditions de vie matérielles et culturelles des familles. Nous avons ainsi choisi de réaliser l'enquête qualitative dans les XVIII^e, XIX^e et XX^e arrondissements de Paris, et ce avant même de choisir ces mêmes arrondissements comme terrain de l'expérimentation, car ces quartiers se caractérisent par une forte mixité sociale⁴³. Ce terrain d'enquête ouvrait la possibilité de réaliser des entretiens avec des parents de situation très modeste, typiquement des femmes seules, avec ou sans emploi, éventuellement avec de nombreux enfants, n'ayant pas fait d'études en France ou ayant eu un parcours scolaire difficile, mais aussi avec des parents de situation plus caractéristique des classes moyennes, voire moyennes et supérieures, typiquement celles de couples de cadres et/ou professions intellectuelles, avec un ou deux enfants.

L'enquête a permis de rassembler 21 entretiens, représentant un total de plus de 12 heures d'enregistrements. Les entretiens, de type semi-directif, ont été conduits sur la base de guides d'entretien comprenant trois parties : une première centrée sur l'environnement familial de l'enfant, une deuxième sur l'organisation de son temps extrascolaire et une troisième sur les pratiques de lecture parentale. Certes, ces entretiens ont été effectués avant le tirage au sort des écoles et n'ont donc pas été réalisés (ou seulement par hasard) avec des parents par la suite impliqués dans le programme. Ils ne fournissent donc pas des données réelles mais seulement virtuelles sur sa mise en œuvre et sa réception. Mais les données dont nous disposons n'en portent pas moins sur des familles dont l'expérience quotidienne est proche de celles qui ont bénéficié du programme, et chaque entretien était utilisé pour présenter le programme que

⁴¹ Michael Patton, *Qualitative Evaluation and Research Methods*, Londres, Sage, 1990.

⁴² A. Revillard, « Saisir les conséquences... », art. cité, p. 473.

⁴³ Yann Caenen *et al.*, « Une mosaïque sociale propre à Paris », *Analyses*, INSEE Île-de-France, 53, février 2017.

nous souhaitons mettre en place, pour recueillir leur opinion sur cette initiative. Ces entretiens constituent de ce fait des sources d'information permettant de saisir le « point de vue » des publics ciblés par l'expérimentation : s'entretenir directement avec les parents permettait de simuler avec eux une reconstitution du quotidien de leur vie de famille et offrait à ce titre l'occasion de recueillir des données de type factuel, détaillées et circonstanciées, sur les pratiques de lecture avec leurs enfants, de manière à pouvoir imaginer différents types de mécanismes activés par le programme que nous souhaitons expérimenter.

Ces entretiens ont permis de documenter une grande diversité d'expériences quotidiennes de la lecture parentale. Celles-ci apparaissent en premier lieu dépendre fortement des ressources, au sens large du terme (culturelles, sociales, économiques), dont disposent les parents⁴⁴. Parmi ces ressources, une attention particulière a été accordée à l'offre de livres à domicile. Les entretiens ont permis d'observer que la variété de cette offre selon les familles, en quantité et qualité, est liée à la conscience plus ou moins vive des parents de l'intérêt de renouveler, de diversifier et d'ajuster le contenu des livres en fonction de l'âge et du développement cognitif de leur enfant.

L'expérience quotidienne de la lecture partagée dépend en second lieu de la variable temporelle et d'obstacles tels que le travail de nuit ou le stress professionnel pour parvenir à organiser un moment consacré à la lecture entre parents et enfants. Ces obstacles sont naturellement susceptibles d'entraver la réceptivité pratique des familles à un programme tel que nous l'avons testé. Un programme aura beau distiller des informations et des conseils aux parents pour les aider à insérer une séquence régulière de lecture à la maison, si leur situation professionnelle (par exemple celle d'un père allophone réceptionniste de nuit ou celle d'une femme célibataire exerçant comme pédiatre urgentiste) perturbe, voire rend impossible, la conciliation de leur rythme de travail avec l'emploi du temps des enfants, sa mise en œuvre ne pourra avoir qu'un impact très limité.

Les données recueillies ont également permis de documenter une grande diversité de rapports à l'institution scolaire. Ceux-ci peuvent aller des relations les plus dépendantes, et en même temps les plus confiantes, aux situations où les parents expriment peu d'attentes à l'égard de l'école tout en pouvant adhérer aux valeurs qu'ils associent à l'école publique, en passant par toute une gamme de rapports plus subis, en raison des contraintes économiques pesant sur

⁴⁴ Cela a été récemment documenté et analysé de manière approfondie pour le contexte français dans Marianne Woollven *et al.*, « Le langage comme capital », dans Bernard Lahire (dir.), *Enfances de classes. De l'inégalité parmi les enfants*, Paris, Seuil, 2020, p. 1037-1042.

le lieu de résidence, ou de rapport plus directement critiques, que ce soit en raison du traumatisme que leur a laissé une expérience scolaire difficile ou d'un parcours migratoire nourri d'attentes à l'égard de l'école qui n'ont pas été satisfaites. Ce rapport à l'institution scolaire paraît difficilement dissociable de la variété des littératies familiales, c'est-à-dire de la variété des rapports culturels entretenus à l'égard de l'écrit, des livres, et plus largement de la langue française, dans la mesure où celles-ci varient selon les individus et leur milieu social d'appartenance. Il convient notamment de ne pas occulter le caractère potentiellement ambigu des relations qu'entretiennent les catégories populaires avec l'institution scolaire⁴⁵, cela nécessitant d'être appréhendé au regard de la problématique de la réception de l'action publique.

Mieux comprendre la réception de l'action publique par l'élaboration d'idéaux-types

Dans une perspective d'évaluation des politiques publiques, la « réception »⁴⁶ désigne un processus qui « combine des effets et des appropriations de l'action publique »⁴⁷ dont l'analyse consiste à prendre en compte les dimensions habilitante et contraignante du social, ainsi que les dimensions objective (les pratiques) et subjective (les représentations) du comportement des destinataires. L'étude de la réception de l'action publique, à dominante qualitative, est ainsi une démarche utile pour enrichir, en les contextualisant, les résultats d'une recherche ayant adopté une méthode à dominante quantitative. Si, en elles-mêmes, les techniques qualitatives de collecte de données permettent de documenter des situations seulement individuelles et singulières, leur analyse par interprétation compréhensive permet d'en dégager des paramètres généraux et la construction d'idéaux-types⁴⁸. Dans le cas du projet « *Language Skills* », l'étude de la réception a consisté à construire des idéaux-types de parents permettant d'enrichir un raisonnement de type causal, centré sur des variables, avec les résultats

⁴⁵ Pour une approche de cette question dans une perspective de sociologie de la culture voir Claude Grignon, Jean-Claude Passeron, *Le savant et le populaire. Misérabilisme et populisme en sociologie et en littérature*, Paris, Gallimard, 1989, p. 83-84 ; Pour une synthèse récente des travaux de sociologie de l'éducation conduits sur ce thème depuis les années 1990 en France, voir Lorenzo Barrault-Stella *et al.*, « Les rapports à l'école comme rapports à l'État », *Politix*, 130, 2020, p. 7-22, ici p. 10-11.

⁴⁶ Philippe Warin, « La réception des politiques par les destinataires. Esquisse d'une approche des relations entre action publique et lien politique », note de synthèse pour l'habilitation à diriger des recherches en science politique, Grenoble, université Pierre-Mendès-France, 1998 ; A. Revillard, « Saisir l'action publique... », art. cité.

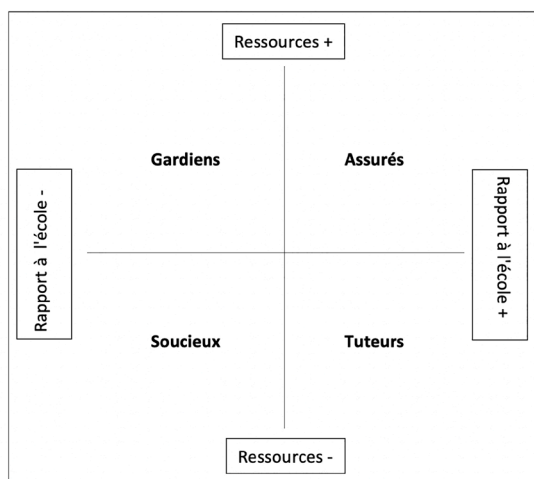
⁴⁷ A. Revillard, « Saisir l'action publique... », art. cité, p. 479.

⁴⁸ Max Weber, *Essais sur la théorie de la science*, Paris, Plon, 1965, p. 180 *sq.* ; Jean-Claude Passeron, *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*, Paris, Nathan, 1991.

propres à une approche compréhensive, centrée sur l'étude de cas⁴⁹.

Nous avons construit notre typologie des parents en reprenant les deux dimensions constitutives de la variété des expériences quotidiennes de lecture parentale (ressources et rapports à l'institution scolaire) et en les utilisant comme des axes polarisant les situations des parents : l'axe des ressources va des expériences de lecture les plus exposées à la précarité à celles qui le sont le moins⁵⁰ ; l'axe sur les rapports à l'école s'ordonne autour de la plus ou moins grande familiarité avec la culture scolaire, estimée par le niveau de formation des parents, tout en intégrant la dimension potentiellement conflictuelle du rapport des milieux populaires à l'école. Le croisement de ces deux axes permet de dégager quatre idéaux-types de parents (voir figure 2).

Figure 2. Typologie des parents



Les parents « soucieux » correspondent à un type de parents au niveau de précarité élevé, disposant au plus d'un diplôme du secondaire, mais entretenant une relation de dépendance confiante envers l'institution scolaire pour ce qui concerne leurs enfants. Tout en adhérant au principe d'un programme de lecture parentale venant de l'école, ces parents ne peuvent en pratique s'y engager de manière suffisamment constante pour que des effets positifs puissent être observés sur les compétences langagières de leurs enfants.

⁴⁹ Dominique Schnapper, *La compréhension sociologique*, Paris, PUF, 2018.

⁵⁰ La construction de cet axe s'inspire du constat de polarisation croissante du monde social entre « protégés » et « assistés », induit par l'évolution des politiques sociales depuis les années 1990 et leurs effets en termes de régulation sociale. Voir sur ce point Bruno Palier, *Gouverner la sécurité sociale. Les réformes du système français de protection sociale depuis 1945*, Paris, PUF, 2005 ; Serge Paugam, Nicolas Duvoux, *La régulation des pauvres*, Paris, PUF, 2008.

Les parents « tuteurs » sont des parents vivant dans des conditions matérielles modestes mais disposant d'un niveau de formation suffisant pour s'investir de manière durable dans les activités pédagogiques à domicile et entretenir des interactions continues avec l'école. En plus d'adhérer sans réserve au principe d'un programme de lecture parentale, ces parents veillent et parviennent à s'organiser en conséquence pour que leurs enfants en tirent bénéfice.

Les parents « gardiens » constituent une figure de parents pouvant vivre dans des conditions matérielles précaires mais qui disposent d'un niveau de formation suffisant pour déployer une stratégie éducative qui les amène à cultiver une distance critique à l'égard de l'institution scolaire. L'implication de ces parents dans un programme de lecture parentale est plus incertaine, de même que l'observation d'un effet du programme sur le développement des compétences langagières de leurs enfants.

Enfin, le type de parents « assurés » désigne des parents vivant dans des conditions matérielles élevées et possédant un niveau de formation suffisamment important pour les rendre confiants à l'égard de la réussite scolaire de leurs enfants sans avoir à s'investir de manière soutenue dans les interactions avec l'institution scolaire. Quand bien même ces parents ne se sentiraient pas hors cible d'un programme de lecture parentale, celui-ci n'est pas de nature à impacter fortement leur quotidien et les compétences langagières déjà élevées de leurs enfants.

Cette typologie fournit des clés de compréhension relatives à la réception, à l'appropriation et aux effets qu'un programme de lecture parentale est susceptible d'avoir. Les deux premiers types, que les observations de terrain conduisent à estimer les plus nombreux, sont prédisposés par principe et/ou en pratique à se saisir d'un tel programme, tandis que les deux autres y prédisposent de manière plus incertaine ou correspondent à des parents dont les caractéristiques sociales, plus favorisées, sont de nature à minimiser l'impact potentiel du programme. On peut ainsi imaginer que les parents « soucieux » auront entendu parler du programme et auront pu échanger quelques mots à son sujet avec leurs enfants mais en ne consultant que de manière distraite les brochures fournies et en n'utilisant qu'occasionnellement les livres. Les parents « tuteurs », sitôt le programme lancé, en auront « joué le jeu » de manière plus méthodique, en utilisant les livres comme un support supplémentaire d'activités avec leurs enfants et en s'appropriant le contenu des brochures. Les parents « gardiens », tout en consentant formellement au programme par souci de ne pas en exclure leurs enfants, n'auront pas cherché à jouer un rôle de relais efficace du programme à la maison, en ne lisant pas ou de manière peu impliquée les brochures ainsi que les livres prêtés. Quant aux parents « assurés », ils auront bien été informés du programme et auront eu à son sujet une opinion positive dans l'intérêt de la plupart des familles fréquentant l'école de leurs

enfants, mais l'auront vu avant tout comme quelque chose venant répliquer ce qu'ils font déjà avec leurs enfants. Une telle reconstitution du comportement des parents aide à comprendre que le programme testé dans le projet « *Language Skills* » est susceptible de bénéficier prioritairement aux enfants des « parents tuteurs ».

Plus largement, la typologie produite aide à identifier de nouveaux enjeux en termes d'action publique et d'évaluation. Un premier enjeu serait de parvenir à ajuster la forme et le contenu du programme testé de manière à augmenter autant que possible la probabilité que les parents « soucieux » et « gardiens » aient envie d'y participer ou parviennent à en tirer davantage profit. Il serait par exemple possible d'explicitier davantage que les grand.e.s sœurs et frères peuvent jouer un rôle déterminant dans l'instauration d'une habitude de lecture partagée et le développement du goût pour l'écrit. On pourrait également orienter davantage les familles vers des supports de lecture autres que les livres prêtés et pour autant facilement mobilisables au quotidien, comme des émissions de télévision ou chaînes YouTube de lecture de livres pour enfants. Un deuxième enjeu tient au fait que le volet qualitatif de la recherche permet de montrer que les inégalités éducatives ont une pluralité de sources qui ne peuvent pas être éliminées uniquement par une intervention de nature informationnelle. L'analyse de la réception suggère ainsi qu'il faudrait concevoir d'autres types de programmes (par exemple des programmes ciblés sur la réduction de la précarité) prenant davantage en compte les contraintes de ces familles.

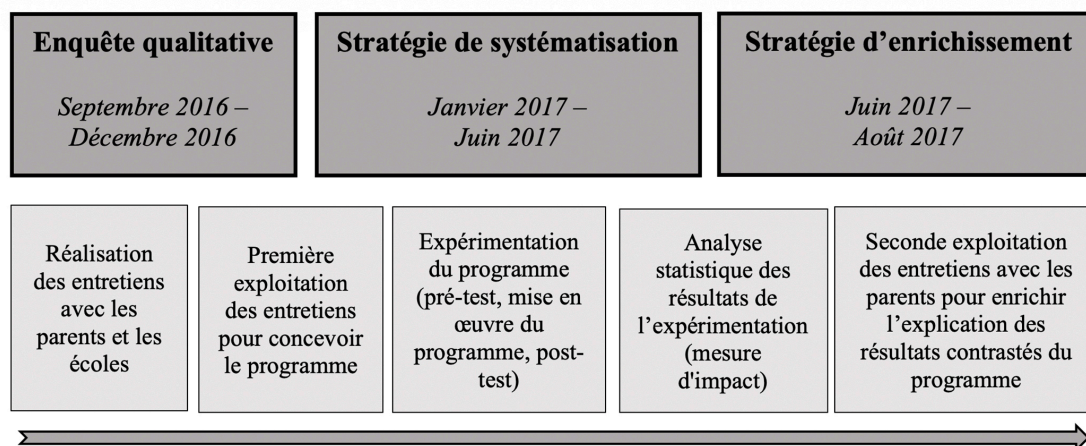
De telles observations ou idées ne peuvent se formuler aussi clairement lorsque l'on se limite à faire un essai randomisé et à évaluer son impact. Le recours aux méthodes mixtes est très utile pour comprendre la variabilité de l'impact d'un programme et pour en apprécier les limites. L'apport du « qualitatif autonomisé », en particulier, est d'aider à concevoir de nouvelles hypothèses, non initialement formulées et donc non testées, mais qui ouvrent des pistes de recherche pour des expérimentations ultérieures. Il contribue à ce titre à nourrir un processus d'apprentissage et de délibération collectifs, sur une base scientifique, portant sur le développement d'un programme d'action et les objectifs qui lui sont assignés.

Conclusion

L'article avait pour objet d'illustrer deux stratégies de méthodes mixtes, celle de systématisation et celle d'enrichissement. Ces deux stratégies sont présentées sur la base d'un projet d'évaluation *ex ante* qui recourait à une démarche expérimentale à dominante quantitative tout en recourant en amont et en aval à une méthodologie qualitative. Cette double

combinaison des méthodes quantitatives et qualitatives peut être schématisée comme ci-dessous.

Figure 3. Schématisation de la double combinaison des méthodes qualitatives et quantitatives dans le projet « *Language Skills* »



La première partie de l'article a présenté la stratégie de recherche globale du projet, afin de rendre compte de la perspective théorique de l'évaluation réaliste dans laquelle ont été combinées les deux manières d'articuler les méthodes quantitative et qualitative. Précisant la stratégie de systématisation, la deuxième partie a souligné que celle-ci ne se réduit pas à l'usage d'une méthodologie seulement quantitative, mais qu'elle intègre, en amont, un travail de recherche qualitatif « instrumentalisé » pour identifier les conditions de transposition d'un protocole expérimental dans des conditions pratiques et locales nécessairement singulières. Enfin, la troisième partie a montré que, dans le cadre d'un projet de recherche adoptant un protocole expérimental, la conduite d'une stratégie d'enrichissement consistait à mobiliser les méthodes qualitatives en aval pour compléter l'identification d'un lien causal entre un programme et ses effets par la compréhension des processus ayant abouti aux impacts mesurés. Une manière de conduire ce type de stratégie est de produire une analyse centrée sur la réception de l'action publique, celle-ci pouvant donner lieu, comme dans le cas du projet « *Language Skills* » à la construction d'idéaux-types. Plus largement, dans le cadre du dossier, ce texte permet de faire ressortir l'intérêt de la typologie des méthodes mixtes présentée dans l'article introductif : la capacité à articuler dans un même projet de recherche deux stratégies de méthodes mixtes sera probablement dorénavant facilitée par le fait de les avoir plus clairement distinguées.

Si l'article avait pour objet principal d'explicitier l'apport des méthodes mixtes pour

l'évaluation d'un programme par expérimentation randomisée, il était aussi l'occasion de formuler une thèse spécifique au regard de la littérature existante sur l'évaluation réaliste. La littérature francophone sur les aspects épistémologiques et méthodologiques de l'évaluation est peu développée, et lorsqu'elle s'y intéresse en référence à la démarche réaliste, c'est soit pour se concentrer sur d'autres méthodes que l'expérimentation, à savoir l'usage de la revue systématique⁵¹, soit, lorsqu'elle s'intéresse à l'expérimentation randomisée, pour en fournir une représentation opposée dans ses principes à une démarche d'évaluation réaliste⁵². Nous souhaitons montrer pour notre part que le recours à l'expérimentation n'était pas antinomique avec une démarche réaliste et même qu'elle pouvait en constituer un précieux auxiliaire à condition de l'envisager dans le cadre de l'expérimentation de terrain et du recours aux méthodes mixtes. L'expérimentation randomisée n'est en soi qu'une technique qui s'insère plus largement dans des options méthodologiques traduisant elles-mêmes des postures épistémologiques.

Le projet « *Language Skills* » ne visait pas à valider ou invalider de manière catégorique l'efficacité des programmes de lecture parentale mais à identifier en quoi, pour qui, comment et à quelles conditions ils pouvaient l'être tout en étant cohérents avec les finalités de réduction des inégalités de réussites scolaires, historiquement présentes dans les politiques éducatives en France, et réactivées à nouveaux frais depuis la publication de grandes études à dominante quantitative, telles que celle du PISA, du Programme international de recherche en lecture scolaire (PIRLS), ou de *Teaching And Learning International Survey* (TALIS), qui permettent de comparer les systèmes éducatifs nationaux. À ce titre, un des apports de l'article est de nourrir la réflexion sur l'*evidence-based policy*, en soulignant l'intérêt d'appréhender les échanges entre recherche et action publique, moins sous l'angle du transfert (direct) de résultats qui émaneraient du monde de la science seule que sous celui de l'apprentissage collectif entre acteurs sociaux, politiques et scientifiques, impliqués dans une même démarche orientée vers le développement d'une évaluation de nature scientifique.

⁵¹ E. Robert, V. Ridde, « L'approche réaliste pour l'évaluation de programmes et la revue systématique... », art. cité.

⁵² A. Devaux-Spatarakis, « L'évaluation "basée sur la théorie"... », art. cité.