



HAL
open science

Légitimation et légitimité des réformes

Denis Meuret

► **To cite this version:**

Denis Meuret. Légitimation et légitimité des réformes. Revue Internationale d'Education de Sèvres, 2020, 83, pp.113-121. 10.4000/ries.9372 . halshs-02896113

HAL Id: halshs-02896113

<https://shs.hal.science/halshs-02896113v1>

Submitted on 10 Jun 2021

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.

Légitimation et légitimité des réformes

Denis Meuret

Université de Bourgogne

Pour toute réforme, on peut distinguer deux phases, celle de sa justification (les arguments qui président à sa conception, à son adoption et aux efforts pour persuader les acteurs de sa validité) et celle de son évaluation. La première a lieu avant et pendant la mise en œuvre de la réforme. La phase de l'évaluation, formelle ou informelle, se place après la mise en œuvre : la réforme tient-elle ses promesses ? Ces deux phases ont été transformées par les progrès considérables, depuis une quarantaine d'années, des bases de données et des moyens de leurs traitements. La recherche a acquis un tel pouvoir de légitimation qu'un procès, actuellement en cours d'instruction, a été intenté par sept élèves contre le système d'éducation du Michigan, au motif qu'il les privait de leur droit constitutionnel à la maîtrise de la lecture, entre autres à cause d'un « manque d'enseignement de la lecture fondé sur des preuves (*evidence based*) » (Superfine *et al.*, 2019). Des évaluations internationales, dont PISA, on peut tirer des indications sur les caractéristiques des systèmes scolaires les plus efficaces et équitables (Schleicher, 2018) ou des établissements scolaires les plus efficaces – c'est-à-dire les plus efficaces à coût égal – (Agasisti et Zoido, 2018). Par ailleurs, il est devenu plus facile de mesurer les résultats d'une politique pour décider de la continuer ou non.

En France, deux politiques récentes ont été justifiées par la recherche : la décision de réduire les redoublements a été prise en s'appuyant sur des recherches qui montraient que les redoublants progressaient moins que les élèves de même niveau qu'on laissait passer dans la classe supérieure. Une recension des nombreuses études américaines sur ce sujet (Crahay, 2003) a joué un grand rôle dans cette décision. D'autre part, la récente politique de réduction de la taille des classes en réseau d'éducation prioritaire (REP) dans les premières années de l'enseignement primaire, doit beaucoup à des recherches françaises (Piketty et Valdenaire, 2006) et américaines (entre autres, Nye *et al.*, 2011) qui convergeaient pour diagnostiquer un effet particulièrement positif de cette réduction sur cette population (élèves socialement défavorisés en début du primaire).

Inversement deux politiques dont les justifications empiriques étaient faibles ont été mises en veilleuse : la « réforme des rythmes scolaires » (2013) et le dispositif « Plus de maîtres que de classes » (PMQC), permettant aux équipes pédagogiques des écoles défavorisées de décider elle-même de l'activité d'un enseignant supplémentaire (2012). Dans les deux cas, on peut noter un déficit de justification et un déficit d'évaluation.

Pourtant, à des périodes où la recherche empirique en éducation était balbutiante, et aujourd'hui encore, de nombreuses réformes ont été décidées sans lien avec cette recherche et sont considérées comme réussies. On se propose ici d'étudier les vertus et les limites de l'approche pragmatique des réformes et la façon dont cette approche peut s'articuler avec d'autres.

JUSTIFICATION

Certaines réformes ne peuvent, par nature, être déduites de la recherche empirique. Ce sont celles qui s'appuient sur l'idée qu'un nouvel état du monde, en train d'advenir, appelle une nouvelle forme d'école. On peut en donner de multiples exemples, anciens ou récents.

Émile Durkheim propose, au tournant du ^{xx}^e siècle, une forme d'école adaptée à un monde où la religion ne discipline plus les corps et les esprits : l'école apprendra l'autodiscipline.

Également au début du ^{xx}^e siècle, John Dewey propose une école adaptée à la démocratie américaine, qui développe le goût de l'expérimentation et du service. Cette conception de l'école influencera fortement l'École mise en place pendant la *Progressive Era* (1890-1920) et permet de comprendre certaines différences actuelles entre l'école française et l'école américaine (Meuret, 2007).

Plus récemment, au cours des années 1960, la commission Parent propose au Québec une école adaptée aux temps nouveaux : moins malthusienne parce que le monde a besoin de travailleurs plus qualifiés, développant un « nouvel humanisme » qui ne se définit plus en opposition à la science et à la technique, où filles et garçons seraient davantage égaux parce que cette inégalité-là ne sera pas longtemps tolérée, etc. Ce rapport irrigue encore aujourd'hui les discours sur l'École au Québec, l'une des meilleures du monde (Meuret, 2013).

Des réformes ont, en Finlande, en Norvège, en Suède dans les années 1960 ou 1970, à Singapour un peu plus tard, bouleversé avec tant de succès des systèmes scolaires centralisés, malthusiens et autoritaires que tout le monde est persuadé aujourd'hui que leur forme actuelle date de toujours et correspond à une « culture » profondément ancrée de leurs peuples.

Encore plus récemment, la Corée du Sud, pourtant parmi les nations en tête du classement PISA, a considéré que son système d'enseignement n'était plus adapté aux besoins d'une économie moderne. Le gouvernement lui-même a critiqué un système scolaire trop axé sur le « par cœur », une pédagogie trop centrée sur l'enseignant, et une trop grande prégnance de l'évaluation des élèves. Cette critique a inspiré plusieurs réformes depuis 2009, visant à développer davantage la créativité des élèves et leur capacité d'autonomie, à faire diminuer la forte pression exercée sur eux par les systèmes d'évaluation. Par exemple, un semestre sans évaluation dans les collèges doit permettre aux élèves de prendre part à un large éventail d'activités et d'expériences (Gim, 2019). Ces réformes coréennes s'inspirent d'ailleurs d'une réforme mise en œuvre au Japon depuis 2002. Elles sont critiquées par certains chercheurs, qui pointent la faiblesse des méthodes proposées pour développer la créativité et craignent une augmentation des inégalités dans l'un des systèmes scolaires les plus équitables de la planète selon PISA (Park, 2015).

Au colloque dont ce numéro rend compte, dans l'atelier qui est le point de départ de cet article, cinq réformes ou projets de réforme ont été présentées¹.

1. Colloque « Conditions de réussite des réformes en éducation » organisé par la *Revue internationale d'éducation de Sèvres*, CIEP, 12-14 juin 2019, atelier 2 « Comprendre la mise en œuvre des politiques éducatives : nouveaux éclairages », coordonné par Bénédicte Robert. [<https://journals.openedition.org/ries/6658>]

Deux étaient justifiées par des recherches empiriques. Gonzalez (2019) vise à favoriser la mise en œuvre, dans les écoles d'Amérique latine, de modes d'enseignement « fondés sur la preuve ». D'une analyse des résultats de la recherche mondiale, les promoteurs ont tiré 33 pratiques pédagogiques adaptées aux conditions de l'enseignement en Amérique latine, pratiques qu'ils ont ensuite essayé de disséminer dans la région. Gurgand (2019) expose comment une expérimentation menée en France dans l'académie de Créteil (« La mallette des parents ») a montré qu'une action simple et peu coûteuse (donner à des parents éloignés de l'école des informations de base sur le fonctionnement du collège et de l'orientation) se traduisait par une diminution sensible de l'absentéisme des élèves et par une amélioration notable de leur orientation, et comment cette expérimentation est ensuite devenue une politique systématique au niveau académique.

Mais les trois autres reposaient sur d'autres types de justification. *Teaching at the Right Level* (TRL), un mode d'enseignement très élaboré, a été proposé en Inde par la fondation Pratham, à partir moins de la recherche que d'une forte dose de bon sens, qui conduisait à chercher une alternative à un enseignement qui se souciait fort peu de savoir si les élèves apprenaient réellement ce qu'on leur exposait (Lakhsman, 2019). Hargreaves (2019) expose un projet de refondation de l'école en Ontario, qui repose sur l'idée que l'évolution du monde soumet l'école à de nouveaux défis, auxquels elle doit répondre par un véritable basculement, celui d'un âge de « l'effort pour réussir » à un âge de « l'apprentissage, de l'identité et du bien-être ». Ce basculement s'opèrera par une réforme radicale de l'évaluation, par l'adoption par les enseignants d'une culture de « professionnalisme collaboratif » et par une gouvernance de l'école non plus depuis le haut mais « depuis le milieu », c'est-à-dire depuis des enseignants autonomes, coopérant entre eux et diffusant en réseaux leurs bonnes pratiques. Annan (2019) a élaboré en Nouvelle-Zélande un projet un peu semblable : le monde actuel, complexe, requiert des élèves engagés plus que des élèves réussissant des tests. Il faut redéfinir ce que nous appelons « succès », en finir avec le gouvernement *top-down* de l'École, impliquer les enseignants dans la fabrication du curriculum et dans la recherche de ce qui marche. Cela favorisera leur engagement, leur enthousiasme et de ce fait leur capacité à produire ces jeunes engagés que réclame le monde qui vient.

On voit que ces deux derniers projets non seulement ne demandent pas à la recherche empirique d'indiquer les réformes à mener, mais font appel aux enseignants et non aux chercheurs, assimilés à une autorité supérieure², pour décider des pratiques à mettre en œuvre.

Les recherches fondées sur la recherche empirique, quant à elles, vont chercher dans des expérimentations de laboratoire ou dans des observations de situations réelles des formes d'enseignement plus efficaces ou plus équitables. Cependant, l'idée que la recherche dicte les politiques à mettre en œuvre correspond à une représentation erronée de la science. Oui, des progrès en nombre considérable ont été enregistrés dans le monde par l'effet de l'application de démarches scientifiques, plutôt qu'idéologiques, à la plupart des problèmes que rencontre l'humanité

2. Qu'il faille assimiler des chercheurs, qui déduisent de leur observation du monde ce qui est le plus efficace, à des individus qui décideraient ce qui marche de façon autoritaire pour l'imposer ensuite aux masses obéissantes est une opinion dont je laisse la responsabilité à ces deux auteurs.

(Pinker, 2018). Mais c'est précisément parce que la science suggère à ces problèmes des réponses révisables, moins assurées que ne le faisaient l'idéologie ou la religion. La science n'est pas une autre source de vérité incontestable, elle procède par expérimentation, par essais et erreurs, n'est jamais certaine (Dewey, 2014). Labrousse (2016) a tort de dénoncer les bricolages auxquels sont contraintes les recherches expérimentales sans dire que les autres types de recherche y sont aussi contraints. La vertu de la science est seulement de ne pas laisser une erreur prospérer longtemps, mais c'est un avantage décisif sur les autres prétendants à la vérité.

Il en résulte qu'il serait désastreux d'ignorer les résultats des recherches empiriques pour décider d'une politique, mais aussi que qui voudrait s'appuyer sur eux comme s'il s'agissait d'une vérité révélée écrasant toute autre approche de son autorité ferait fausse route. En voici, plus concrètement, trois raisons.

1) La science progresse, c'est sa caractéristique. Par exemple, un article récent sur les stratégies d'amélioration des écoles les plus efficaces (Sun *et al.*, 2019), construit une typologie par l'analyse informatique de milliers de stratégies d'amélioration et de rapports sur leur mise en œuvre dans l'État de Washington (États-Unis), puis isole les stratégies qui se traduisent par les plus grandes améliorations pour les élèves. Le progrès par rapport aux premières recherches sur l'amélioration des écoles (des données recueillies, des méthodes d'analyse) dans les années 1990 est impressionnant. Mais ces progrès modifient parfois le jugement que l'on peut porter sur une politique. Par exemple, les recherches sur l'effet de la réduction de la taille des classes ont longtemps conclu à sa minceur, avant qu'une étude expérimentale à large échelle ne vienne changer la donne. Par exemple encore, les recherches sur le redoublement ont été moins sévères avec ce dernier, quand elles ont mieux mesuré l'équivalence des populations comparées, ceux qui passaient dans la classe supérieure et ceux qui redoublaient. Évidemment, ces progrès sont, pour les adversaires de la science, le signe qu'on ne peut lui faire confiance.

2) Le diable est dans les détails. Il est parfois difficile de savoir pourquoi tel dispositif réussit là où un autre, *a priori* identique, échoue ou, ce qui revient au même, de savoir ce qui au juste est évalué. Certains dispositifs sont faciles à cerner, mais des interventions plus complexes ne le sont pas. Ainsi, un dispositif danois proche de « Plus de maîtres que de classes » a fait récemment l'objet d'une évaluation rigoureuse et positive (Andersen *et al.*, 2018), alors que les évaluations précédentes de dispositifs semblables (par exemple, Blatchford, 2012) avaient été négatives.

3) La recherche est parfois impuissante à emporter l'adhésion des acteurs. On a montré aux États-Unis que la mesure de l'efficacité des enseignants par des épreuves passées par leurs élèves en début et en fin d'année scolaire était la plus juste (Bacher-Hicks *et al.*, 2019). Mais on a montré aussi que cette modalité d'évaluation était la plus rejetée par les enseignants (Jiang *et al.*, 2015).

Tout ceci signifie que, si la recherche empirique est devenue un élément nécessaire de la justification des réformes, elle ne peut en être la source unique, en particulier lorsque cette justification repose sur la description d'un monde à venir qui ne peut, par nature, être observé, mais aussi parce que ce qui fait son mérite (sa perfectibilité, sa capacité d'aller contre de fausses évidences) constitue aussi sa faiblesse en tant qu'outil de persuasion.

ÉVALUATION

Lorsqu'une réforme est justifiée par un récit sur les besoins du monde qui vient, et non par les résultats de recherche, il peut arriver que partisans et adversaires s'affrontent sur la base de principes et d'idéologies, mais aussi qu'une expérimentation les départage. Lorsque la Suède, dans les années cinquante, la Norvège un peu plus tard, décidèrent d'allonger la scolarité obligatoire et d'ouvrir l'enseignement secondaire, il n'a pas manqué de personnes dans ces pays pour vouer ces réformes aux gémonies, selon des arguments bien connus : inutilité de doter les travailleurs de compétences élevées, baisse du niveau, etc. De fait, d'ailleurs, il s'agissait bien d'un pari : aucune recherche empirique ne pouvait prouver que le pays avait bel et bien besoin d'une population mieux formée. Dans ces deux pays, le gouvernement fit en sorte que la réforme s'appliquât sur une partie seulement du territoire et on compara le niveau des élèves et les performances économiques dans les deux parties. Le premier et les secondes s'avérant meilleurs dans les zones réformées, la réforme fut étendue à tout le pays (Maurin, 2007).

Mais même les réformes fondées sur la recherche appellent une évaluation. Les meilleures recherches comparent deux populations d'élèves, identiques parce que tirées au sort, dont la politique étudiée a été appliquée à l'une et pas à l'autre (le groupe témoin). Elles peuvent être faites en laboratoire ou *in situ*. Elles s'efforcent de faire en sorte que cette application soit optimale pour que soient vraiment mesurés les effets de cette politique. C'est de bonne méthode, mais cela ouvre la possibilité que les choses se passent différemment sur le terrain. C'est pourquoi l'évaluation est nécessaire, qui indiquera les effets de la politique telle qu'elle est réellement mise en œuvre.

Voici deux exemples qui montrent l'importance de cette question de la mise en œuvre :

- Aux États-Unis, un programme pour l'apprentissage de la lecture a été mis au point à partir des meilleures recherches sur le sujet, puis des milliers d'enseignants ont été formés à le mettre en œuvre. Jacob (2015) a mesuré que les élèves des enseignants ainsi formés n'apprenaient pas mieux la lecture que les autres. Une enquête auprès des enseignants formés lui a appris qu'ils étaient satisfaits de la formation reçue, mais qu'ils n'enseignaient tout simplement pas de la façon indiquée par le programme.
- Le programme indien évoqué ci-dessus, *Teaching at the Right Level* (TRL), a été évalué par l'équipe d'Esther Duflo sur quelques écoles choisies pour étudier la faisabilité du programme. L'évaluation fut très positive et, de ce fait, le programme fut implanté à large échelle. Cette implantation elle-même fut évaluée par la même équipe. Duflo (2018) dit des résultats de cette évaluation : « Verdict : ça n'a pas marché. Pourquoi ? parce que les enseignants, obnubilés par la tyrannie des programmes nationaux, n'y avaient pas recours ». Décision a donc été prise de prévoir un moment et un lieu spécifique à l'enseignement TRL et, en outre, la hiérarchie a insisté sur son soutien au programme. Les enseignants se sont alors sentis autorisés à le mettre vraiment en œuvre et une évaluation a cette fois montré son efficacité, conduisant à l'exporter dans de nombreux pays de sorte qu'il touche aujourd'hui plus de 4 millions d'enfants (Duflo, 2018).

Ces deux exemples montrent que l'évaluation ne peut se limiter à mesurer si une politique est efficace ou non, si une réforme produit ou non les effets attendus. Si elle s'y limite et que les résultats sont décevants, il sera facile aux adversaires de la politique d'incriminer la politique elle-même et à ses partisans d'incriminer une mise en œuvre déficiente. Il faut donc que l'évaluation s'efforce d'expliciter les raisons des résultats obtenus, ce qui peut s'avérer plus difficile que de mesurer les résultats.

Le cas de la politique de l'éducation prioritaire en France est instructif à cet égard. Cette politique a été lancée en France alors que des politiques semblables avaient échoué, dans d'autres pays, à améliorer significativement les apprentissages des élèves défavorisés. Par ailleurs trois évaluations ont conclu de même pour la France (Meuret, 1994 ; Bénabou *et al.*, 2004 ; Davezies *et al.*, 2013), sans que cette politique soit abandonnée. L'histoire de l'évaluation de l'éducation prioritaire en France montre d'abord qu'une politique appuyée sur un principe de justice difficilement récusable (« donner plus à ceux qui ont moins ») peut résister à des évaluations décevantes. Elle montre aussi que le jugement des évaluations est moins indiscutable qu'on ne le croit parfois. Meuret (1994) montrait que, dix ans après la mise en œuvre de la politique, les élèves de ZEP progressaient un peu moins que des élèves hors ZEP de même origine sociale et de même niveau initial. Mais cette étude laissait ouverte la possibilité que la politique des ZEP ait réduit, sans l'annuler, un désavantage dû non pas aux caractéristiques individuelles des élèves, mais à un effet de contexte. Benabou *et al.* (2004) réduisent fortement cette possibilité, en prenant comme critère la différence de performance après et avant la politique entre des couples de collèges semblables, les uns passés en ZEP avec la politique, les autres demeurés hors ZEP. L'étude conclut à un effet nul de la politique et propose, peut-être trop rapidement, que la raison en est la trop faible surdotation des établissements ZEP. Restait cependant la possibilité que cette différence nulle signale en fait une réussite de la politique : elle aurait réussi à compenser une plus forte dégradation du contexte dans les zones prioritaires. Davezies *et al.* (2013) évaluent une politique compensatoire un peu modifiée à la suite de l'évaluation de Bénabou *et al.* (moins d'établissements, avec une plus forte surdotation). Ils tiennent l'évolution du contexte sous contrôle en comparant l'évolution des performances de collèges situés juste de part et d'autre du seuil d'éligibilité à la politique, donc susceptibles d'avoir vu leurs contextes évoluer de façon semblable. Cette étude, comme la précédente, trouve un effet nul de l'éducation compensatoire : que le collègue bénéficie ou non de la politique ne change rien pour la réussite scolaire des élèves. Sa limite est cependant de ne comparer avec les collèges hors éducation prioritaire que les collèges en éducation prioritaire situés dans les zones les moins difficiles, elle ne peut exclure que la politique marche mieux pour les collèges en situation plus difficile. Il est possible que, si les évaluations avaient conclu à un résultat négatif plutôt que nul, la politique aurait été remise en cause. Toujours est-il qu'elle ne l'a pas été, entre autres parce que les évaluations, bien que rigoureuses, n'ont pas paru assez probantes pour que le risque soit pris de déroger au principe de justice cité plus haut. Si on avait pu montrer qu'une autre politique, visant aussi à diminuer les inégalités sociales de performance scolaire, obtenait des résultats plus probants³, il en aurait peut-être été différemment.

3. C'est le cas de deux politiques anglaises (*Literacy Hour* et *Excellence in the Cities*).

L'évaluation est aussi contrainte par le fait que l'évaluation ne tombe pas, si l'on peut dire, dans des oreilles seulement soucieuses de rigueur scientifique. Le débat à propos des évaluations d'une réforme suppose un minimum d'exigence et de bonne foi, à la fois lors de la réalisation et de la réception de l'évaluation. Si l'on traite de propagande ministérielle les résultats d'une évaluation rigoureuse qui s'avère positive, si l'on traite d'idéologie les résultats des évaluations qui ne confirment pas sa propre idéologie, on mine les présupposés même de l'utilité du processus d'évaluation. Par ailleurs, de bonne foi, les adversaires d'une réforme pourront mettre en avant ses possibles effets pervers, potentiellement très nombreux. Ainsi la régulation par les résultats mise en place aux États-Unis, en 2003, a été accusée de limiter l'apprentissage à la capacité de passer les tests, de stresser les élèves, de dégoûter les enseignants du métier, de concentrer l'enseignement sur les matières testées et sur les élèves moyens faibles au détriment des plus faibles, de favoriser la triche. Ces critiques étaient d'ailleurs soutenues par une certaine version du récit deweyen, contre laquelle la recherche ne pouvait lutter que besogneusement, en étudiant sérieusement la validité de chaque critique (Meuret, 2012). Ceci montre que le processus d'évaluation peut être complexe, coûteux, et que sa vertu pour la prise de décision est conditionnée à un certain niveau de rigueur et de transparence de la part des évaluateurs, mais aussi de bonne foi de la part des évalués.

MODESTE PROPOSITION POUR UN NOUVEAU RÉCIT SUR L'ÉCOLE

Ce qui précède conduit à l'idée que le rôle accru de la recherche empirique dans la justification des réformes doit être pensé avec, et non contre, celui de récits globaux sur ce que l'école doit faire pour le pays. Ces récits peuvent aller là où la recherche ne peut aller (l'avenir), et leur partage par les acteurs se traduit par une réception moins conflictuelle, moins soupçonneuse des résultats de la recherche ou des méthodes de l'évaluation. Ils indiquent des orientations quant aux réformes souhaitables, mais aussi des questions à la recherche ainsi que des objets et des critères à l'évaluation. Plus généralement, ils contribuent à faire de l'enseignement une profession attirante et favorisent ainsi la qualité des enseignants, qualité qui est une des caractéristiques principales des systèmes d'enseignement les plus efficaces et équitables selon PISA (Schleicher, 2018).

En se référant à l'ampleur des défis auxquels ont su faire face les récits qui ont réussi à réformer l'école en profondeur, il vient qu'un défi équivalent aujourd'hui est la menace qui pèse, non pas sur la planète, mais sur nos sociétés à cause du réchauffement climatique. Le prévenir autant que possible, vivre avec la température que l'on n'aura pas su éviter sont deux enjeux que l'École peut aider à affronter. À bien y réfléchir, d'ailleurs, il est étrange qu'on nous serine que nous devons changer radicalement notre façon de produire, de consommer, d'être cause du réchauffement, sans qu'on ait essayé d'imaginer le rôle que pourrait jouer l'École dans ce changement. Il en est encore temps : si l'on suppose qu'il faudrait cinq ans pour qu'un nouveau récit sur l'école soit élaboré, diffusé et fasse sentir ses effets, celui-ci touchera les générations dont la scolarité obligatoire se déroulera entre 2025 et 2040. Or on estime que la période de plein impact du réchauffement commencera d'ici 2040 à 2050⁴.

4. Cf. Alexandre Magnan, chercheur du GIEC (Groupe international d'études sur le climat), interview sur le site *Good Planet Mag*, le 25 septembre 2019. [<https://bit.ly/2j3Dwem>]

Un tel récit devrait définir les qualités dont la nouvelle École devrait doter les élèves (rigueur pour penser le diagnostic, créativité pour imaginer des solutions, empathie vis-à-vis des victimes, résilience face aux catastrophes) ou qu'elle devrait présenter elle-même (par exemple, équité pour générer un sentiment de solidarité). La recherche empirique pourrait ensuite jouer un grand rôle en aidant à discerner les compétences scolaires et les politiques les mieux à même de développer ces qualités.

BIBLIOGRAPHIE

AGASISTI T. et ZOIDO P. (2018). "Comparing the efficiency of schools through international benchmarking: Results from an empirical analysis of OECD PISA 2012 data". [<https://doi.org/10.3102/0013189X18777495>]

ANNAN B. (2019). « Facteurs de réussite de la mise en œuvre de la politique éducative néozélandaise ». *Revue internationale d'éducation de Sèvres*, colloque *Conditions de réussite des réformes en éducation*, CIEP, 12-14 juin 2019. [<https://journals.openedition.org/ries/7332>]

ANDERSEN S.C., BEUCHERT L., SKYT NIELSEN H. et KJRGAAARD THOMSEN M. (2018). "The effects of teachers aides in the classroom: evidence from a randomized trial". *Journal of the European Economic Association*, 00(0)1-38. [<https://doi.org/10.1093/jeaa/jvy048>]

BACHER-HICKS A., CHIN M., KANE T. et STEIGER D., (2019). "An experimental Evaluation of three teachers quality measures: Value added, classroom observation, students review". *Economics of Education Review*, 73. [<https://doi.org/10.1016/j.econedurev.2019.101919>]

BÉNABOU R., KRAMARZ F. et PROST C. (2004). « Zones d'éducation prioritaire, quels moyens pour quels résultats ? Une évaluation sur la période 1982-1992 ». *Économie et statistiques*, n° 380, p.3-34. [DOI : <https://doi.org/10.3406/estat.2004.7676>]

BLATCHFORD P., RUSSELL A. et WEBSTER R. (2012). *Reassessing the impact of teaching assistants: How research challenges practice and policy*, Oxon : Routledge.

CRAHAY M. (2003). *Peut-on lutter contre l'échec scolaire ?* (2^e édition). Louvain-la-Neuve : De Boeck.

DAVEZIES L. et BEFFY M. (2013). « Has the "Ambition Success Networks" Educational Program Achieved its Ambition? ». *Annals of Economics and Statistics*, GENES, n° 112, p. 271-293.

DEWEY J. (2011). *Démocratie et éducation*, Paris : Armand Colin (édition originale, 1916).

DEWEY J. (2014). *La quête de la certitude*, Paris : Gallimard.

DUFLO E. (2018). « Très peu d'expériences sont réalisées en France ». Interview par Florence Rosier, *Le Monde*, 20 février.

GIM C. (2019). « Efficience et gouvernance : l'exemple de la Corée du Sud ». *Revue internationale d'éducation de Sèvres*, colloque 2019 « Conditions de réussite des réformes en éducation », CIEP, 12-14 juin. [<http://journals.openedition.org/ries/7731>]

GONZALEZ J. (2019). « De l'évidence mondiale à sa transposition dans les pratiques locales, l'expérience latino-américaine ». *Revue internationale d'éducation de Sèvres*, colloque « Conditions de réussite des réformes en éducation », CIEP, 12-14 juin. [<https://journals.openedition.org/ries/7362>]

- GURGAND M. (2019). « Interventions scolaires expérimentales dans des conditions écologiques. *Revue internationale d'éducation de Sèvres*, colloque « Conditions de réussite des réformes en éducation », CIEP, 12-14 juin. [<https://journals.openedition.org/ries/7402>]
- HARGREAVES A. (2019). « Collaborative Professionalism and Leading from the middle in an era of complex policy change ». *Revue internationale d'éducation de Sèvres*, colloque « Conditions de réussite des réformes en éducation », CIEP, 12-14 juin. [<https://journals.openedition.org/ries/7490>]
- JACOB B. (2017). « When evidence is not enough, findings from a randomized evaluation of evidence based literacy instruction (EBLI) ». *Labour Economics*, vol 45 (C), p. 5-16.
- JIANG J.Y., SPORTE S.E. et LUPPESCU S. (2015). « Teacher perspectives on evaluation reform: Chicago's REACH students ». *Educational Researcher*, 44 (2), p. 105-116. [<https://doi.org/10.3102%2F0013189X15575517>]
- LABROUSSE A., (2016). « Apprendre des expérimentations aléatoires ». *La vie des idées*, [<https://bit.ly/3a7foU2>]
- LAKHSMAN S. (2019). « Améliorer les résultats en lecture et en arithmétique à grande échelle : « Enseigner au bon niveau » (*Teaching At The Right Level*), l'approche de Pratham ». *Revue internationale d'éducation de Sèvres*, colloque « Conditions de réussite des réformes en éducation », CIEP, 12-14 juin. [<https://journals.openedition.org/ries/7470>]
- MAURIN E. (2007). *La nouvelle question scolaire*. Paris : Le Seuil (Points).
- MEURET D. (1994). « L'efficacité de la politique des zones d'éducation prioritaire dans les collèges », *Revue française de pédagogie*, n° 109, p. 41-64. [<https://doi.org/10.3406/rfp.1994.1245>]
- MEURET D. (2007). *Gouverner l'École*, Presses Universitaires de France.
- MEURET D. (2012). « Les effets de l'*accountability* sur les politiques d'éducation aux États-Unis ». *Éducation et sociétés*, 30, p. 75-87. [<https://bit.ly/2WvdRDj>]
- MEURET D. (2013). *Pour une école qui aime le monde*. Presses Universitaires de Rennes.
- NYE B., HEDGES L.V. et KONSTANTOPOULOS S. (2000). « The effects of small classes on academic achievement: the results of the Tennessee class size experiment ». *American Educational Research Journal*, 37(3), p. 123-152.
- PARK H. (2014). « Très Bien mais peut mieux faire, les réponses apportées par le système éducatif coréen aux résultats de l'enquête PISA ». *Revue internationale d'éducation de Sèvres*. [<https://journals.openedition.org/ries/3727>]
- PIKETTY T. et VALDENNAIRE M. (2006). « L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées ». *Dossiers Education et formation* n° 173. [<https://bit.ly/3djm7fU>]
- PINKER S. (2018). *Le triomphe des Lumières*. Paris : Les Arènes.
- SCHLEICHER A. (2018). *World Class: How to build a 21st-century school system, Strong Performers and Successful Reformers in Education*, Paris : OECD Publishing.
- SUN M., LIU J., ZHU J., LE CLAIR Z., (2019). « Using a text as data approach to understand Reform Processes: A deep exploration of school improvement strategies ». *Educational Evaluation and Policy Analysis*, 41 (4), p. 510-536.
- SUPERFINE B.M., GOLDMAN S. et RICHARD M.S. (2019). « Toward a synergetic model for improving the use of research in court-driven educational reform : Examining Gary B. vs Snyder and literacy Improvement in Detroit ». *Educational Researcher*, 48 (8), p. 543-548. [<https://doi.org/10.3102/0013189X19874067>]