



Peut-on vraiment évaluer les politiques d'éducation prioritaire en Europe

Marc Demeuse

► To cite this version:

Marc Demeuse. Peut-on vraiment évaluer les politiques d'éducation prioritaire en Europe. Marc Demeuse, Daniel Frandji, David Greger, Jean-Yves Rochex. Les politiques d'éducation prioritaire en Europe. Vol 2 : questions sur le devenir et les possibles en matière d'égalité scolaire, ENS éditionsN, pp.281-292, 2011, 9782847883220. hal-00763952

HAL Id: hal-00763952

<https://hal.science/hal-00763952>

Submitted on 8 Jan 2013

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.

Peut-on vraiment évaluer les politiques d'éducation prioritaire en Europe ?

Marc Demeuse

Institut d'Administration Scolaire, Université de Mons, Belgique

Introduction

Pour répondre à la question introduite par le titre de ce chapitre, il faut envisager à la fois l'angle technique (dispose-t-on les moyens d'évaluer ce type de politiques ?) et l'angle de l'autorité et des usages (s'autorise-t-on à évaluer ce type de politiques ? que fait-on de ces évaluations?). C'est ce que feront respectivement les auteurs de deux chapitres qui vont suivre. Le premier chapitre (Demeuse, Demierbe et Friant) proposera de clarifier l'analyse, suggérant un modèle global qui sera appliqué à certaines des PEP qui ont été analysées dans le premier ouvrage (Demeuse et al. 2008). Le second chapitre (Broccolichi) interrogera les mésusages et les lacunes dans l'évaluation de ces mêmes politiques.

Dans ce chapitre introductif, nous tenterons de mieux définir ce que recouvre le concept d'évaluation et d'apporter quelques précisions à propos des méthodes d'évaluation dans le contexte des politiques d'éducation prioritaire.

Ce qu'évaluer veut dire

Il existe plusieurs modèles d'évaluation¹ et ce terme recouvre donc plusieurs sens. « Les politiques publiques sont traditionnellement évaluées selon trois approches temporelles, une approche *ex-ante* qui dresse le bilan de la situation sectorielle et permet d'identifier le référentiel de la politique à mettre en œuvre en fondant les décisions prises, une approche par le suivi qui permet de cerner les obstacles et/ou facteurs facilitateurs de mise en œuvre des décisions, et enfin une approche *ex-post* qui a pour objet de mesurer l'efficacité des décisions prises et celle de la conduite des opérations (l'évaluation *ex-post* pouvant dans certains cas acquérir le statut d'évaluation *ex-ante* si le politique décide d'agir sans délai sur la base de l'évaluation d'une politique antérieurement conduite) » (Solaux, 1999, p. 182). L'approche adoptée dépend, entre autres, de l'objet à évaluer, de l'objectif qui est assigné à cette évaluation, du type d'informations qu'il est possible de recueillir et de l'utilisation qui en sera faite. Ainsi, on mobilisera d'autres démarches lorsqu'il s'agira d'évaluer la pertinence du ciblage d'une politique d'éducation prioritaire (les moyens sont-ils bien accordés aux établissements qui devraient en bénéficier, au regard des objectifs poursuivis ?) ou d'évaluer par exemple l'efficacité d'un programme d'intervention précoce sur la maîtrise de la lecture à l'entrée de l'enseignement obligatoire.

Le terme « évaluation » auquel les auteurs de cette partie de l'ouvrage se réfèrent doit être entendu au sens donné par Stufflebeam et ses collègues (1980, cité par Legendre, 2005, p. 630) d'un « processus par lequel on délimite, obtient et fournit des informations utiles permettant de juger des décisions possibles ». Il s'agit en particulier d'évaluations de programme² dans le cadre de politiques publiques.

¹ Voir par exemple Smith, 1997 pour une discussion du terme « modèle », ou de celui de « théorie », dans le contexte de l'évaluation.

² Voir par exemple, Nadeau (1988) et le volume 29, n°3 (2006) de la revue « Mesure et évaluation en éducation » publiée par l'Association pour le Développement des Méthodologies de l'Evaluation en Education (ADMEE), dont notamment les articles de Figari et Tourmen sur la référentialisation, ou d'Hurteau et ses collègues. Ces derniers auteurs (p. 30) définissent l'évaluation de programme « comme une activité scientifique visant à établir la valeur ou le mérite d'un programme ou d'un de ses éléments constitutifs (House, 1980), et ce, au moyen de déclarations et de conclusions qui sont précises, justifiables,

Ces évaluations peuvent prendre différentes formes et impliquer différents acteurs aussi bien lors de la définition de leur objet et de leurs objectifs, que lors de leur mise en œuvre³. Il peut ainsi s'agir d'évaluations externes⁴, d'évaluations internes⁵ et même d'évaluations mixtes négociées⁶ dans lesquelles, par exemple, des équipes de recherche reçoivent une commande, mais jouissent de la liberté nécessaire pour mener à bien cette mission, compte tenu des règles de l'art. Dans cette dernière configuration, sans doute la plus complexe, il s'agira de bien clarifier le rôle et les responsabilités de chacun. Ainsi, il faudra être très attentif au(x) rôle(s) que le commanditaire pourrait souhaiter consciemment ou non voir jouer par l'équipe de recherche en gardant à l'esprit la distinction essentielle proposée par Nadeau (1988, p. 38) entre évaluation et recherche. Ces deux démarches sont entreprises pour des raisons différentes : « la curiosité scientifique serait l'élément motivant du chercheur alors que la contribution à la solution d'un problème particulier constituerait la motivation de l'évaluateur ». Pour autant, comme le signale Blais (2006, p. 130), la volonté de reddition de compte ou *accountability* très présente, notamment à travers l'élaboration d'indicateurs destinés à évaluer l'efficacité de telle ou telle politique, entraîne dans son sillage « la recherche 'évaluative' et la pousse à s'éloigner de la simple description et du jugement pour explorer le terrain miné de la recherche des causes de l'efficacité ou de l'inefficacité »⁷.

Si les PEP peuvent connaître différents âges, il semble bien que les rôles potentiels des chercheurs et leurs relations avec les commanditaires connaissent aussi des « périodes types » : du rôle d'observateur et d'analyste extérieur à celui de soutien à la mise en œuvre des politiques elles-mêmes à un niveau local ou national (comme, par exemple, en Grèce ou en Belgique) en passant par celui de « conseiller en prospective », chargé d'identifier ce qui est susceptible de marcher ou de soutenir la réflexion politique (Aubert-Lotarski et al. 2007 ; Demeuse, Derobertmeasure & Friant, 2010).

Dans la suite de cette partie, nous ne traiterons que des évaluations dites « formelles », au sens où il s'agit d'une démarche « systématique qui s'appuie sur des données obtenues méthodiquement à l'aide d'instruments et de grilles qui se focalisent sur des indicateurs d'observation et qui disposent de critères précis d'interprétation » (Legendre, 2005, p. 645). Dans ce contexte, les instruments dont il est question dans la définition de Legendre peuvent prendre différentes formes (observations directes en classe, questionnaires, exploitations de documents ou de données recueillies par d'autres) et le traitement des données n'aboutit pas nécessairement à une quantification.

Il va de soi que la démarche adoptée devra toujours présenter les gages nécessaires pour pouvoir être prise au sérieux et nous présenterons quelques-unes de ces conditions dans la suite de ce texte. Cependant, comme le souligne Dauvisis (2006, p. 58), « démarche de preuve – ou au moins souvent présentée comme telle –, l'évaluation est en réalité une pratique sociale contextualisée, régie par des normes, des coutumes et des valeurs qui ne s'inscrivent pas toutes dans la logique d'une totale rationalité. Comme toute pratique sociale, elle n'échappe pas à des jeux d'acteurs nombreux, et l'implicite y règne de manière évidente permettant aux différents acteurs de détenir une part de pouvoir à préserver jalousement : toute tentative de clarification est

acceptables, légitimes ; bref, qui sont valides (Fournier & Smith, 1993). » Il importe aussi que « la conclusion de cette démarche soit argumentée par une information dite suffisante pour être acceptée du client de l'évaluation. Les conclusions évaluatives pourront alors servir à la prise de décision concernant la programmation future du programme ».

³ En matière d'évaluation des politiques éducatives, on ne se dispensera pas trop rapidement de la vérification de la mise en œuvre effective d'un programme (contrôle de conformité) de manière à s'assurer que les décisions prises ont bien été appliquées avant de tenter d'en estimer les résultats.

⁴ « Evaluation conduite par un évaluateur qui provient de l'extérieur de l'organisation où prend place l'étude » (Nadeau, 1988, p. 419).

⁵ « Evaluation conduite par un membre de l'organisation où l'étude prend place » (Nadeau, 1988, p. 419).

⁶ Voir aussi Aubert-Lotarski et ses collègues (2006) à propos de la conduite des audits à visée participative.

⁷ Nisbet (1997, pp. 211-212) précise que la « policy-oriented research » est habituellement définie par opposition à la recherche fondamentale et fait l'objet de différentes qualifications : « appliquée » (par opposition à « fondamentale »), « policy-oriented » (par opposition à « curiosity-oriented »), « à fonction instrumentale » (par opposition à « enlightenment function »), « tournée vers la décision ou l'action » (par opposition à « orientée vers la connaissance ou la théorie ») et même « pertinente » (« relevant » en langue anglaise) (par opposition à « académique »). Par delà les appellations plus ou moins flatteuses, ce sont les objectifs, l'autonomie des chercheurs et le caractère commandité ou en réponse à une demande (par opposition à une recherche autonome, y compris dans la définition de ses questions) qui semblent caractériser ce type de recherche, en limitant parfois la portée générale.

perçue avec un risque de mise en péril d'un équilibre social toujours fragile. » Le chapitre de Broccolichi abordera en particulier les évaluations des PEP comme pratiques sociales à travers les mésusages des démarches évaluatives et de leurs lacunes.

Attention ! Méthode(s)

L'évaluation de programme repose sur une approche méthodique et demande une solide formation. Si les méthodes et les modèles ont évolué⁸, tout comme les techniques⁹, on peut identifier plusieurs approches susceptibles de fournir des éléments utiles en matière d'évaluation de programme.

Approches expérimentale, quasi-expérimentale et corrélationnelle

Ce n'est que grâce à une expérimentation menée avec soin qu'il est possible de s'assurer véritablement de l'efficacité d'un programme, notamment en s'assurant de comparaisons avec des groupes de contrôle et une assignation aléatoire des élèves (qui n'est pas présente dans les études quasi-expérimentales¹⁰). Une véritable expérimentation contrôlée, telle que décrite par exemple dans le « manuel de psychopédagogie expérimentale » de Léon et ses collègues (1977) et appliquée au domaine de l'évaluation de programme ne se rencontre guère dans le monde latin. Cette approche n'est pourtant pas rare, notamment aux Etats-Unis.

Il est vrai, dans cet environnement, que les enjeux financiers liés à la démonstration de l'efficacité d'un programme particulier sont importants : les fonds fédéraux attribués au titre du *Title I* sont en effet majorés lorsque les programmes impliquent des établissements scolaires de manière globale (et non à travers un programme ciblé sur l'une ou l'autre matière) et qu'ils ont montré clairement leur efficacité (Borman et al. 2001).

Cela amène de nombreux chercheurs à s'efforcer de démontrer l'efficacité des programmes qu'ils développent, comme par exemple dans le cas de *Success for All* de la *Johns Hopkins University*. Ainsi, par exemple, Robert Slavin, le promoteur de *Success for All* (Slavin et al. 1996), rédige-t-il de nombreuses publications à propos des résultats de son programme (principalement en recourant à des modèles quasi-expérimentaux avec groupes de contrôle) et va-t-il jusqu'à proposer une analyse comparative d'autres programmes dans un ouvrage au titre évocateur (*Show Me the Evidence. Proven and Promising Programs for America's Schools*, Slavin et Fashola, 1998). Il ne s'agit pas d'un cas unique: Wang, Haertel et Walberg (1998), promoteurs d'un autre programme, proposent « un guide comparatif » dans la revue *Educational Leadership*. Ces textes font l'objet de nombreuses discussions méthodologiques dans les revues scientifiques nord-américaines.

Il n'est pas inintéressant de constater que de l'autre côté de l'Atlantique, si les programmes destinés aux écoles accueillant des élèves « à risque » sont l'objet d'enjeux financiers très importants – ces écoles pouvant adhérer à l'un ou l'autre des programmes –, cela pousse les chercheurs universitaires à la fois à développer des procédures rigoureuses de contrôle de la qualité et de nombreuses publications critiques dans des revues de très haut niveau¹¹. On ne peut malheureusement pas en dire autant dans les pays où les programmes sont développés en dehors de tout support académique, soit que les chercheurs s'impliquent peu, soit qu'ils sont tenus à l'écart par les structures politico-administratives en charge des programmes.

A côté des études expérimentales et quasi-expérimentales, « les études "ex post", qui sont menées de manière rétrospective, si elles peuvent l'être avec beaucoup de rigueur, ne permettent pas de manipuler ni de

⁸ Voir par exemple l'ouvrage, en langue française, de Nadeau (1988) à propos de l'évolution des modèles et des méthodes, principalement dans le monde anglo-saxon.

⁹ On peut à la fois souligner l'évolution des techniques statistiques, comme l'introduction des analyses multi-niveaux dans la modélisation statistique (Bressoux, 2008) et le recours aux modèles de mesure, et l'évolution des moyens techniques de recueil de données (par exemple, mise à disposition de larges bases de données, utilisation de la vidéo et du codage des comportements en temps réel).

¹⁰ L'étude de Meuret (1994) constitue un bon exemple de comparaison de deux groupes (élèves en ZEP et élèves hors ZEP), sans assignation des élèves de manière aléatoire à chacun des groupes.

¹¹ Pour une illustration des débats autour du programme *Success for All* de Robert Slavin, on peut consulter l'article de Demeuse et Denooz (2001).

contrôler les différentes variables ou situations. Elles ne présentent donc pas la même puissance que les études expérimentales. Le manque de contrôle et de prise d'informations fiables, plus que l'exception, constitue véritablement la règle dans le domaine des réformes éducatives et curriculaires dans les pays européens de langue française » (Demeuse et Strauven, 2006, p. 164).

L'approche corrélationnelle¹², si elle est fréquente, ne permet pourtant pas de s'assurer des facteurs qui influencent, au sens causal véritable, les résultats observés. Cela ne la prive cependant pas de tout intérêt comme le reconnaît Crahay (2005, p. 96) dans son plaidoyer pour une complémentarité des méthodes expérimentales et écologiques en éducation : le caractère parfois artificiel de l'expérimentation peut conduire à détruire l'objet que l'on se propose d'observer. C'est notamment ce que pointe Bronfenbrenner (1979, p. 19), comme le rappelle Crahay, lorsqu'il écrit que « l'essentiel de la psychologie du développement, telle qu'elle existe actuellement, est une science des comportements exceptionnels produits par des enfants placés dans des situations artificielles en présence d'adultes étrangers pendant des périodes de temps les plus brèves possibles ». Plus qu'une question de méthode, c'est la nature des liens entre variables qui est problématique : inférer un lien causal alors que seules des corrélations sont calculées, voilà l'erreur. Ce faisant, parler de l'efficacité d'un programme dans une approche corrélationnelle n'est donc pas adéquat. Cela n'empêche pas, comme nous le verrons par la suite, que d'autres aspects intéressants peuvent être étudiés et pour lesquels une approche corrélationnelle n'est pas inadéquate.

Approche comparative

L'approche comparative qui a été adoptée par l'équipe EuroPEP permet de mettre en perspective les spécificités et les aspects communs des politiques étudiées qui ne peuvent apparaître qu'en plaçant chaque système en regard des autres, après avoir déterminé une grille commune d'analyse et en tentant d'éviter les pièges signalés par Frandji (2008) dans l'introduction du premier ouvrage publié par cette équipe. Cette approche permet également, pour une part, de dépasser la controverse entre les tenants et les adversaires d'une politique donnée en les amenant sur le terrain des faits et non plus seulement sur celui des idées, comme le suggère Sowell (2004, p. 1) dans un ouvrage qu'il consacre à un sujet proche du nôtre, l'« *Affirmative Action Around the World* »¹³.

Mons (2007a et b), tout en soulignant les apports importants des comparaisons internationales, malgré leurs conclusions encore limitées, précise utilement les améliorations nécessaires dans ce type de démarches comparatives, au-delà de leur composante quantitative :

¹² L'approche corrélationnelle consiste à recueillir des informations relatives à de nombreuses variables sans que la situation ne puisse être manipulée expérimentalement et, par le calcul d'indices d'association statistique (corrélation), à établir des relations entre ces variables, sans qu'il soit possible de s'assurer de liens causaux entre celles-ci. Les études internationales en matière d'évaluation - PISA en est certainement l'exemple le plus emblématique - fournissent ainsi un grand nombre de données qu'il est possible de mettre en rapport en terme de co-évolution, sans pour autant pouvoir s'assurer que c'est effectivement la fluctuation d'un paramètre qui entraîne l'évolution d'autres variables dans un sens donné, comme seules les études expérimentales randomisées peuvent le permettre. Il peut en effet exister des causes communes à deux ensembles de variables et il est parfois difficile de dégager le sens de la causalité. On peut citer, comme exemple trivial dans le premier cas, la corrélation qui existe, durant la scolarité obligatoire, entre la pointure des chaussures des élèves et leur niveau en mathématique... Ce n'est néanmoins pas parce que les élèves ont de plus grands pieds qu'ils sont plus compétents en mathématique... c'est simplement parce qu'en fonction de l'avancée en âge le pied s'allonge et le niveau scolaire augmente ! (la corrélation disparaîtra d'ailleurs à l'âge adulte) Quant au sens de la liaison, il peut être assez difficile de savoir si c'est parce qu'ils sont plus motivés que les élèves réussissent mieux en classe ou s'ils sont plus motivés à étudier parce qu'ils réussissent mieux en classe... sans doute s'agit-il d'un processus interactif qu'indique dans ce cas l'indice de corrélation qui unit généralement ces deux variables (motivation et réussite scolaires).

¹³ « Alors que la controverse fait rage aux Etats-Unis à propos de la politique d'*Affirmative Action*, peu d'Américains semblent connaître l'existence ou la pertinence de politiques similaires dans d'autres pays autour du monde. Au contraire, les arguments en faveur ou contre cette politique tendent de part et d'autre à invoquer l'histoire et les traditions qui distinguent les Etats-Unis. La préférence accordée à certains groupes et les quotas existent dans d'autres pays qui ont connu des histoires et des traditions complètement différentes et, dans certains pays, ces politiques existent depuis bien plus longtemps qu'aux Etats-Unis. Que pouvons-nous apprendre de ces expériences ? Existe-t-il des similitudes, des raisons communes, des résultats semblables ? Ou bien la situation américaine est-elle unique ? De manière ironique, la revendication ou la croyance en une spécificité nationale est l'une des caractéristiques les mieux partagées par de nombreux états qui pratiquent ce type de préférence ou de quotas, quel qu'en soit le nom. »

- Le niveau d'analyse ne doit pas être exclusivement le niveau de l'État-nation, celui-ci ne correspondant par exemple pas au niveau de décision en matière d'éducation dans la plupart des pays connaissant une structure fédérale (Allemagne, Belgique, Suisse, États-Unis d'Amérique...).
- La complexité des politiques éducatives doit davantage être capturée par des indicateurs *ad hoc* de nature typologique que par une analyse unidimensionnelle.
- La création de concepts communs s'impose au-delà des catégorisations nationales (ce qui permet d'éviter sans doute une part de « naturalisation » qui s'imposerait par l'évidence de pratiques trop exclusivement contextualisées).
- Le développement d'une approche néo-institutionnelle¹⁴, ne se limitant pas à une analyse formelle et exclusivement descriptive, est nécessaire.
- Les performances des systèmes scolaires doivent être appréciées sous différents angles et pas seulement sous le seul angle de l'efficacité. L'équité, mais aussi la diversité des performances au regard du territoire ou l'importance numérique de certains groupes (par exemple, les plus « forts » ou, au contraire, les plus « faibles »), doivent pouvoir être pris en compte, aussi bien dans le domaine cognitif que non cognitif.
- Les résultats doivent être interprétés dans une perspective relativiste, privilégiant une forme de « généralisation contextualisée » plutôt que la recherche de règles immuables et indépendantes des conditions nationales et locales¹⁵, sans pour autant verser dans l'incomparabilité et le particularisme absolu qui dénierait à l'approche comparative tout intérêt en dehors de l'établissement de « monographies exotiques ».
- le recours à des données factuelles est essentiel. On ne peut pas se limiter à des « avis d'experts indigènes » qui seraient juxtaposés.

Naturellement, l'approche comparative doit reposer sur des données et les traiter de manière adéquate. On retrouve alors les mêmes problèmes que ceux évoqués précédemment : la nature des liens entre variables et la manière de mettre ces liens en évidence.

Conclusion

Nous avons tenté, dans ce rapide chapitre introductif à la partie dévolue à l'évaluation des PEP, d'attirer l'attention sur certains problèmes posés précisément par leur évaluation. Nous avons principalement souligné l'importance du choix des méthodes et les limites qu'imposent ces choix, notamment parce que, pour cerner de manière précise des phénomènes complexes, certaines simplifications et hypothèses instrumentales sous-jacentes sont nécessaires (Mingat, 1999). Il importe de pouvoir les expliciter, en définir les implications. Cette réflexion est d'autant plus utile que les concepts de « bonnes pratiques » et d'« *evidence based policy* » sont très fréquemment mobilisés dans le contexte européen, en particulier dans le cadre de comparaisons internationales.

S'il ne fait aucun doute que la politique éducative devrait pouvoir s'appuyer sur la recherche en éducation – comme la médecine peut s'appuyer sur les connaissances en biologie ou physiologie – il est par contre beaucoup plus délicat d'identifier les connaissances avérées et les conditions dans lesquelles elles peuvent s'appliquer effectivement. Nous l'avons vu, l'expérimentation en éducation est une activité délicate, à la fois parce qu'elle pose des problèmes éthiques, mais aussi parce qu'elle peut parfois détruire l'objet même de l'étude. Il convient donc d'interroger ce qui est appelé « bonnes pratiques » avec beaucoup de rigueur. Souvent en effet, celles-ci sont davantage le résultat de déclarations que le fruit d'une démarche visant à s'assurer des preuves nécessaires à cette labellisation.

¹⁴ Comme Gazibo et Jenson (2004, p. 203) le précisent, en reprenant les mots de Hall et Taylor (1997) qui identifient trois écoles de pensée distinctes au sein de cette approche, les perspectives néo-institutionnalistes « cherchent toutes à élucider le rôle joué par les institutions dans la détermination des résultats sociaux et politiques »; et de poursuivre: « Toutes ces perspectives placent les institutions au centre de l'analyse, en tant que variables dépendantes (phénomène à expliquer), mais surtout en tant que variables indépendantes (facteurs explicatifs) ».

¹⁵ Reprenant à son compte la conclusion d'Epstein (1988) qui réfute le lien de nécessité entre les positions épistémologiques (particularisme vs positivisme) et les méthodes utilisées (quantitatif vs qualitatif), Mons (2007a, p. 423) prône la complémentarité des méthodes aux différents stades de chaque recherche.

Références bibliographiques

- AUBERT-LOTARSKI, A., LECOINTE, M., MAËS, B., REBINGUET, M. & SAINT-JEAN, M. (2006). *Conduire un audit à visée participative. Etablissements scolaires, services, associations...* Lyon : Chronique sociale et Poitiers : Ecole supérieure de l'Education nationale.
- AUBERT-LOTARSKI, A., DEMEUSE, M., DEROBERTMASURE, A., FRIANT, N. (2007). « Conseiller le politique : des évaluations commanditées à la prospective en éducation. L'évolution des relations entre la recherche et le politique en Communauté française de Belgique. » *Les Dossiers des Sciences de l'Education*, 18, 121-130.
- BLAIS, J.G. (2006). « La mesure pour la recherche et/ou le contrôle en éducation. » In G. Figari & L. Mottier Lopez. *Recherche sur l'évaluation en éducation*. Paris: L'Harmattan, Coll. "Evaluer". 124-131
- BRONFENBRENNER, U. (1979). *The Ecology of Human Development. Experiments by nature and design*. Cambridge: Harvard University Press.
- BORMAN, G.D., STRINGFIELD, S.C., SLAVIN, R.E. (eds.) (2001). *Title I. Compensatory Education at the Crossroads*. Mahwah (NJ) and London: Lawrence Erlbaum Associates.
- BRESSOUX, P. (2008). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles : De Boeck université, coll. « Méthodes en sciences humaines ».
- CRAHAY, M. (2005). « Pourquoi l'expérimentation en éducation ? » In M. Demeuse, A. Baye, M.H. Straeten, J. Nicaise, A. Matoul. *Vers une école juste et efficace. 26 contributions sur les systèmes d'enseignement et de formation*. Bruxelles : De Boeck université, coll. « Economie, Société, Région ». 95-114.
- DAUVISIS, M.C. (2006). L'instrumentation en évaluation. *Mesure et Evaluation en Education*, 29(1), 45-66.
- DEMEUSE, M., DENOOZ, R. (2001). « De l'accroissement de l'efficacité des pratiques éducatives : le cas du programme 'Success for All' mis en œuvre par Robert Slavin et l'équipe de la Johns Hopkins University ». *Cahiers du Service de Pédagogie expérimentale – Université de Liège*, 7-8, 103-127.
- DEMEUSE M., & STRAUVEN C. (avec la collaboration de ROEGIER X.) (2006). *Développer un curriculum d'enseignement ou de formation. Des options politiques au pilotage*. Bruxelles: De Boeck Université.
- DEMEUSE M., DEROBERTMASURE A., & FRIANT N. (2010). « Differentiated financing of schools in French-speaking Belgium: perspectives for regulating a school quasi-market ». *Educational Research and Evaluation: An International Journal on Theory and Practice*, Volume 16, Issue 2, 197 – 216
- EPSTEIN, E. (1988). « The Problematic Meaning of 'Comparison' in Comparative Education ». In J. Schriewer & B. Holmes (eds.). *Theories and Methods in Comparative Education*. Frankfurt: Verlag Peter Lang.
- FIGARI, G. & TOURMEN, C. (2006). « La référentialisation : une façon de modéliser l'évaluation de programme, entre théorie et pratique. Vers une comparaison des approches au Québec et en France ». *Mesure et Evaluation en Education*, 29(3), 2-25.
- FRANDJI D. (2008). « Pour une comparaison des politiques d'éducation prioritaire en Europe ». In M. Demeuse, D. Frandji, D. Greger, & J-Y. Rochex (Eds). *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en oeuvre, débats*. Lyon : INRP, 9-34
- GAZIBO, M. & JENSON, J. (2004). *La politique comparée. Fondements, enjeux et approches théoriques*. Montréal : Les Presses de l'Université de Montréal.

- HALL, P.A. & TAYLOR, R. (1997). « La science politique et les trois néo-institutionnalismes ». *Revue française de science politique*, 47(3-4), 469-496.
- HURTEAU, M., LACHAPELLE, G., HOULE, S. (2006). « Comprendre les pratiques évaluatives afin de les améliorer : la modélisation du processus spécifique à l'évaluation de programme ». *Mesure et Evaluation en Education*, 29(3), 27-44.
- LEGENDRE, R. (2005, 3^e édition). *Dictionnaire actuel de l'éducation*. Montréal (Québec) : Guérin éditeur.
- LÉON, A., CAMBON, J., LUMBROSO, M., WINNYKAMEN, F. (1977). *Manuel de psychopédagogie expérimentale*. Paris: Presses universitaires de France, coll. « Pédagogie d'aujourd'hui ».
- MEURET D. (1994). « L'efficacité de la politique des zones d'éducation prioritaire », *Revue française de pédagogie*, 109, 41-64
- MINGAT A. (1999). « Evaluation des politiques publiques en éducation : quelques balises pour leur conduite. In G. Solaux (dir). *L'évaluation des politiques d'éducation* ». Dijon : CRDP de Bourgogne, coll. « Documents, actes et rapports pour l'éducation ».
- MONS, N. (2007a). « L'évaluation des politiques éducatives. Apports, limites et nécessaire renouvellement des enquêtes internationales sur les acquis des élèves ». *Revue Internationale de Politique comparée*, 14(3), 409-423.
- MONS, N. (2007). *Les nouvelles politiques éducatives. La France fait-elle les bons choix ?* Paris : Presses universitaires de France, Coll. « Education et société ».
- NADEAU, M.A. (1988, 2^e édition). *L'évaluation de programme. Théorie et pratique*. Québec : Les presses de l'Université Laval.
- NISBET, J. (1997). Policy-oriented Research. In J.P. Keeves (ed.). *Educational Research Methodology, and Measurement: An International Handbook*. Cambridge: Pergamon.
- SLAVIN, R.E., FASHOLA, O.S. (1998). *Show me The Evidence! Proven and Promising Programs for America's Schools*. Thousand Oaks (CA): Corwin Press.
- SLAVIN, R.E., MADDEN, N.A., DOLAN, L.J., WASIK, B.A (1996). *Every Child, Every School. Success for All*. Thousand Oaks (CA): Corwin Press.
- SMITH, N.L. (1997). «Evaluation Models and Approaches». In J.P. Keeves (ed.). *Educational Research Methodology, and Measurement: An International Handbook*. Cambridge: Pergamon.
- SOLAUX, G. (1999). « Conclusion ». In G. Solaux (dir.). *L'évaluation des politiques d'éducation*. Dijon : CRDP de Bourgogne, coll. « Documents, actes et rapports pour l'éducation ».
- SOWEL, T. (2004). *Affirmative Action Around the World. An Empirical Study*. New Haven & London : Yale University Press.
- WANG, M.C., HAERTEL, G.D., WALBERG, H.J. (1998). «Models of Reform: A Comparative Guide». *Educational Leadership*, 55(7), 66-71.