

Évaluer les politiques d'éducation prioritaire en Europe: un défi méthodologique

Marc Demeuse et Nathanaël Friant

L'évaluation des politiques éducatives n'est pas une entreprise aisée, à la fois parce qu'elle se heurte à des difficultés méthodologiques importantes, mais aussi parce qu'elle menace ou semble menacer certains projets construits essentiellement sur une base idéologique, alors même que leurs promoteurs s'en défendent le plus souvent. Il est pourtant essentiel de pouvoir évaluer ces démarches politiques, d'autant plus qu'elles concernent des publics défavorisés. Cet article, alimenté par une recherche menée au niveau européen dans huit pays (EuroPEP), met en évidence cette nécessité et illustre à la fois les problèmes rencontrés et les pistes qui peuvent être mises en œuvre. Il indique la nécessité d'une réflexion méthodologique globale et interroge, notamment, la manière dont sont définies «les bonnes pratiques», souvent invoquées dans le domaine des politiques éducatives.

Introduction

Dans la plupart des systèmes éducatifs européens, des politiques volontaristes sont mises en place de manière à rendre l'école plus équitable, le plus souvent en offrant un traitement différencié aux élèves ayant le moins de chances de réussite au départ. Il s'agit de politiques d'éducation prioritaire (PEP). Comme ce texte s'inspire des travaux menés par les auteurs dans le cadre d'une étude européenne intitulée «EuroPEP» (Demeuse, Frandji, Greger & Rochex, 2008, 2011), les politiques d'éducation prioritaires (PEP) dont il sera ici question seront définies comme «des politiques visant à agir sur un désavantage scolaire à travers des dispositifs ou des programmes d'action ciblés (que ce ciblage soit opéré sur des critères ou des découpages socio-économiques, ethniques, linguistiques ou religieux, territoriaux, ou scolaires), en proposant de donner aux populations ainsi déterminées quelque chose de plus (ou de 'mieux' ou de 'différent')» (Frandji, 2008, p. 12).

Puisqu'il s'agit de mettre en œuvre une politique et d'agir sur un désavantage scolaire, une question se pose inévitablement: quels sont les effets des actions

entreprises? C'est là qu'intervient l'évaluation. Cette dernière devrait en effet permettre de fournir, au moment voulu, des informations pertinentes, fiables et objectives sur le rendement du programme ou de la politique, grâce à des méthodes valides et fiables de collecte et d'analyse de données.

Le terme «évaluation» auquel les auteurs de cet article 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 programme* dans le cadre de politiques publiques. Comme le précise Mingat (1999), il existe au moins deux bonnes raisons pour lesquelles il faut évaluer les politiques éducatives: d'une part la complexité de l'organisation éducative, rendant tout à fait possible les erreurs de jugement des décideurs, et d'autre part l'absence de mécanisme interne à l'éducation qui régulerait les erreurs éventuellement commises. L'évaluation pourrait dans ce cas «constituer une sorte de substitut extérieur pour pallier l'absence de mécanismes régulateurs internes. Elle pourrait aussi être à terme intégrée dans les processus ordinaires de gestion des systèmes éducatifs» (p. 35).

L'évaluation systématique des politiques publiques, spécifiquement dans le champ de l'éducation prioritaire, est encore assez rare alors même que ces politiques peuvent être déjà fort anciennes. De fait, l'évaluation de telles politiques n'est pas chose aisée, comme toute évaluation de politiques publiques en matière d'éducation, *a fortiori* lorsqu'il s'agit de politiques concernant des populations réputées défavorisées et qui rompent avec la stricte égalité de traitement. Il faut à la fois l'envisager sous l'angle technique (dispose-t-on des moyens pour évaluer ce type de politiques?) et sous l'angle de l'autorité et des usages (s'autorise-t-on à évaluer ce type de politiques? Que fait-on de ces évaluations?). Cet article privilégie le premier angle de manière à limiter son propos à l'espace qui lui est dévolu.

Évaluer des politiques publiques: de quoi parle-t-on exactement?

Il existe plusieurs modèles d'évaluation¹ et ce terme recouvre donc plusieurs sens. Solaux (1999, p. 182) distingue trois approches temporelles. L'approche *ex-ante* «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». L'approche *par le suivi* «permet de cerner les obstacles et/ou facteurs facilitateurs de mise en œuvre des décisions». Enfin, l'approche *ex-post* «a pour objet de mesurer l'efficacité des décisions prises et celle de la conduite des opérations».

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.

Ces évaluations peuvent donc 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 œuvre. Il peut ainsi s'agir d'évaluations externes², d'évaluations internes³ et même d'évaluations mixtes négociées 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, en effet, 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»⁴.

Si les PEP peuvent connaître différents âges (Rochex, 2008), il semble bien que le rôle 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 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, Demeuse, Derobertmasure & Friant, 2007; Demeuse, Derobertmasure & Friant, 2010).

Il va de soi que la démarche adoptée devra toujours présenter les gages nécessaires pour pouvoir être prise au sérieux. Quelques-unes de ces conditions seront envisagées dans la suite de ce texte. Cependant, comme le souligne Dauvisis (2006, p. 58), l'évaluation ne peut être séparée du contexte social dans lequel elle a lieu:

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 perçue avec un risque de mise en péril d'un équilibre social toujours fragile.

L'*évaluation de programme* repose sur une approche méthodique et demande une solide formation. Si les méthodes et les modèles ont évolué⁵, tout comme les techniques⁶, on peut identifier plusieurs approches susceptibles de fournir des éléments utiles en matière d'évaluation de programme.

Choisir une méthode: approches expérimentale, quasi-expérimentale, corrélacionnelle ou comparative

Approches expérimentale et quasi-expérimentale

Ce n'est que grâce à une expérimentation menée avec soin qu'il est possible d'évaluer véritablement l'efficacité d'un programme, c'est à dire le fait que c'est bien le programme qui produit le résultat observé (et attendu) (Demeuse, Matoul, Schillings & Denooz, 2005), 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-expérimentales⁷). Une véritable expérimentation contrôlée, telle que décrite par exemple dans le «manuel de psychopédagogie expérimentale» de Léon, Cambon, Lumbroso et Winnykamen (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 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 niveau⁸. 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.

Les études quasi-expérimentales ou *ex post* sont menées uniquement de manière rétrospective. Si elles peuvent être très rigoureuses, elles ne présentent toutefois pas la même puissance que les études expérimentales (Demeuse & Strauven, 2006). Elles ne permettent en effet pas de manipuler, ni de contrôler, les différentes variables ou situations. Il faut cependant remarquer, avec Demeuse et Strauven (2006) que «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»

(p. 164).

Approche corrélacionnelle

L'approche corrélacionnelle, si elle est relativement 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), 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» (Bronfenbrenner, 1979, p. 19).

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 corrélacionnelle 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 corrélacionnelle 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 pièges⁹ 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, *l'Affirmative Action Around the World*¹⁰.

Mons (2007a, 2007b), 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:

- 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 na-

tionales (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-institutionnelle¹¹, 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 prises en compte, aussi bien dans le domaine cognitif que dans le domaine 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 locales¹², 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.

L'approche comparative doit donc reposer sur des données et les traiter de manière adéquate. Ce faisant, 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.

Quelques applications pratiques et la nécessité de disposer d'un modèle

EuroPEP a analysé les politiques d'éducation prioritaire dans huit pays européens. Ce projet a donné lieu à plusieurs publications. La première reprend, dans un ouvrage de plus de 450 pages, les analyses pays par pays (Demeuse et al., 2008). Un second ouvrage (Demeuse, Frandji, Greger & Rochex, 2011) propose des analyses transversales de ces politiques en abordant différents aspects, dont le ciblage des populations, les différentes modalités de mise en œuvre (approches préventives et préscolarisation, adaptations/transformations curriculaires) et l'évaluation de ces politiques. Le contenu de cet article est issu de cette troisième partie et s'appuie sur les données récoltées par les différentes équipes nationales et le travail de coordination international des auteurs (Demeuse, Demierbe & Friant, 2011).

En réalité, la question de l'évaluation d'un dispositif, et des politiques d'éducation prioritaire en particulier, ne se limite pas à la seule question de leur efficacité et mérite d'être développée en prenant en compte toutes les facettes. De manière au moins intuitive, on peut formuler d'autres questions en rapport avec

l'évaluation de ces politiques: l'une en rapport avec les actions effectivement menées (que fait-on exactement?) et l'autre en rapport avec le ciblage effectué (cible-t-on bien les élèves qui en ont besoin?). Il est donc nécessaire de pouvoir formaliser les différentes facettes de l'évaluation au moyen d'un modèle.

Bouchard et Plante (2002) proposent une approche analytique de l'évaluation de la «qualité» d'un organisme ou d'une action de formation déclinée en neuf qualités transversales précisément définies: la pertinence, la cohérence, l'à-propos, l'efficacité, la durabilité, l'efficience, la synergie, l'impact et la flexibilité.

Selon les deux auteurs québécois, un programme – par exemple, une politique particulière d'éducation prioritaire – est caractérisé par ces dimensions: «Un ensemble intégré et structuré d'objectifs, de moyens et de personnes, construit en vue de répondre à des besoins qui lui sont externes, assujetti aux contraintes qui lui sont imposées, et redevable des effets, voulus ou non, qui lui sont attribués» (pp. 226-227).

Chaque qualité transversale se définit comme un rapport de conformité entre deux dimensions: ces qualités transversales sont donc évaluées en examinant le degré de correspondance entre deux dimensions (par exemple, l'efficacité est évaluée en examinant le rapport de conformité entre les objectifs et les effets produits par le programme). Complémentaires, les évaluations de chacune des qualités transversales sont amenées à se compléter au sein d'un modèle multidimensionnel. Le recours à des indicateurs permet d'opérationnaliser l'analyse et de sortir d'une impression globale.

Pertinence et à propos

La pertinence est l'une des qualités transversales qui impose une définition précise et opérationnelle des objectifs de la politique considérée et des besoins des individus, particulièrement dans le cas des politiques ciblées, ou du système. C'est évidemment la première qualité d'un programme: les intentions et les objectifs doivent répondre à des besoins clairement identifiés. On comprend dès lors bien les réserves de Frandji (2011) dans l'analyse qu'il fait du programme CLAIR-ÉCLAIR en France¹³.

Évaluer l'à-propos d'un programme, ou de l'une de ses composantes, consiste à examiner dans quelle mesure les contraintes liées au contexte et au terrain ont été prises en compte dans l'élaboration de la politique étudiée. Est évalué le lien de conformité entre la politique et les contraintes de terrain auxquelles elle doit faire face.

Les questions de pertinence et d'à-propos dans les politiques d'éducation prioritaire ont notamment trait à leur ciblage. De telles évaluations sont nombreuses dans les pays ayant participé à l'étude EuroPEP, qu'elles soient faites avant la mise en œuvre de cette politique (par exemple, pour la Belgique francophone, Demeuse & Monseur, 1999) ou après sa mise en application, dans le cas du programme de soutien scolaire en Grèce (Maratou-Alibranti, Teperoglou & Tsiganou, 2006). Par

contre, des informations concernant l'à-propos des PEP sont plutôt fournies par des études prenant en compte le contexte local, par exemple au niveau des zones et des établissements, c'est-à-dire des études menées à une échelle plus réduite, telle que l'étude de Demeuse, Crépin, Jehin et Matoul (2006).

Efficacité, efficacité et impact

A côté de ces deux premiers critères, Bouchard et Plante (2002, p. 230) définissent de manière classique le concept d'efficacité comme

le lien de conformité entre les objectifs visés par [un programme] et les résultats effectivement obtenus. Elle peut être considérée comme l'expression du degré d'atteinte des objectifs réellement visés ou encore comme le nombre d'objectifs effectivement atteints parmi l'ensemble des objectifs effectivement visés.

Leur définition de l'efficacité est, elle aussi, très classique. L'efficacité d'un programme revient à envisager son optimisation, une fois son efficacité avérée. Les auteurs précisent:

Pour être considéré comme efficace, un organisme ou une institution éducative doit satisfaire à l'une ou l'autre des deux conditions suivantes:

- augmenter le degré d'atteinte des objectifs visés sans accroître les moyens alloués dans le fonctionnement pour y parvenir;
- diminuer les moyens alloués pour son fonctionnement, sans diminuer le degré d'atteinte des objectifs visés, c'est-à-dire, sans réduire son efficacité.

(Bouchard & Plante, 2002, p. 231)

Dans certains domaines, un programme peut avoir des résultats qui n'étaient pas prévus et qui ne correspondent pas à des objectifs poursuivis officiellement. Il peut s'agir de l'influence d'un curriculum caché ou, plus largement, de l'effet inattendu de certaines composantes du programme. On peut définir, de manière générale, un critère qui recouvre cette idée et que les auteurs québécois baptisent 'l'impact' et définissent comme

le lien de conformité entre les résultats attribuables à l'organisme, mais non voulus ou non visés de façon explicite à travers ses objectifs, et les exigences sociales, économiques, sociétales, physiques, psychologiques et autres de l'environnement dans lequel il agit et interagit. L'impact prend ainsi en considération la nature des effets non prévus de l'organisme en lien de conformité avec les attentes des divers environnements avec lesquels il est en contact. (Bouchard & Plante, 2002, p. 232)

L'impact peut aussi bien être positif que négatif, impliquer uniquement les élèves ou d'autres individus. Il peut aussi conduire à une remise en cause du

programme lui-même, des objectifs visés...

Quand peut-on dire que l'évaluation d'une politique d'éducation prioritaire nous renseigne sur ses effets? L'évaluation de l'efficacité nécessite de disposer d'une mesure d'entrée de l'atteinte des objectifs et d'une mesure de sortie, réalisées de manière à pouvoir imputer à la PEP les éventuelles évolutions, c'est-à-dire d'un modèle au moins quasi-expérimental (un modèle expérimental étant impossible dans la quasi-totalité des cas). L'évaluation de l'efficience demande d'ajouter à ces éléments la composante «coût» (y compris non monétaire) de la politique. Un exemple d'évaluation de l'efficacité et de l'efficience peut être trouvé dans l'évaluation de la politique d'encadrement renforcé en Communauté flamande de Belgique (Bollens et al., 1998).

Enfin, l'évaluation de l'impact pose davantage de problèmes méthodologiques, puisque cette qualité transversale se centre sur des effets non voulus, et donc ne faisant pas l'objet, en principe, d'une mesure précise, ni à l'entrée, ni à la sortie. L'évaluation ne se basera pas sur des modèles de type quasi-expérimentaux et sera par conséquent moins robuste dans ses inférences causales. Un exemple d'évaluation de l'impact peut être trouvé dans une recherche sur les effets de système de la politique de discrimination positive en Communauté française de Belgique (Bouchat, Delvaux & Hindryckx, 2005, 2008) posant, comme question centrale, l'éventuelle stigmatisation des écoles ciblées par cette politique.

Si comme le montrent Bouchard et Plante (2002), la qualité perçue (ce que les usagers pensent du programme ou du dispositif) est un élément important, à côté de la qualité voulue (annoncée dans les objectifs) et de la qualité rendue (ce qui se déroule réellement, les effets tangibles), les auteurs nous mettent en garde contre une utilisation excessive des évaluations de l'efficacité perçue, souvent présentées à tort comme de réelles évaluations de l'efficacité.

En effet, l'atteinte des objectifs et la satisfaction des usagers, si elles peuvent être liées, ne vont pas automatiquement de pair. Il n'est cependant pas rare que l'on passe de l'évaluation de la satisfaction des usagers à des décisions quant à leur efficacité ou leur efficience, comme cela fut le cas en Grèce lors de la supervision de programmes pilotes (Maratou-Alibranti et al., 2006).

Cohérence et synergie

L'évaluation de la cohérence correspond à l'examen des ressources mobilisées afin d'atteindre les objectifs de la politique prioritaire. Ces ressources regroupent les moyens humains et matériels mis en œuvre. Si les intentions pédagogiques sont bel et bien pertinentes, il convient encore de s'assurer de l'adéquation des moyens et de la compatibilité de ceux-ci entre eux. Par exemple, les enseignants ou les formateurs ont-ils des qualifications adéquates et complémentaires? Le matériel qui leur est fourni peut-il leur être utile pour atteindre les objectifs de l'enseignement ou de la formation? Les instructions qui sont fournies à chacun sont-elles compatibles? Chacun partage-t-il bien les valeurs et les objectifs communs? Voilà quelques questions qui permettent de s'assurer concrètement de la cohérence d'un programme.

La synergie évalue la mise en place et la coordination des moyens humains. Le fonctionnement en place est-il optimal pour obtenir les résultats attendus? Cette démarche se fait à plusieurs niveaux. Ce sont l'action collective de l'équipe et sa dynamique qui sont évaluées. À la prise en compte des aspects économiques des moyens, la synergie ajoute la prise en compte des actions et des caractéristiques des personnes, comme condition susceptible de concourir à l'efficacité. La synergie concerne la coordination, les communications, les attitudes, le climat organisationnel bref, tout ce qui relève de la nature humaine et des interactions entre acteurs.

L'évaluation de la cohérence et de la synergie d'une PEP se produit généralement dans le cadre d'évaluations plus larges intégrant également les qualités transversales que sont l'efficacité ou l'efficacité. Ces deux qualités transversales ne seraient donc pas évaluées pour elles-mêmes, mais plutôt dans le but de mieux identifier les causes de l'inefficacité (avérée ou supposée) d'une PEP, avec pour optique d'en tirer des leçons pour une prochaine réforme. L'évaluation des *Education Action Zones* anglaises en fournit un bon exemple (Antonioni, Dyson & Raffo, 2008; Office for standards in education [Ofsted], 2003; Power, Whitty, Gewirtz, Halpin & Dickson, 2004).

Durabilité et flexibilité

Bouchard et Plante (2002, p. 230), définissent la durabilité comme «le lien de conformité entre les objectifs visés [...] et le maintien, dans le temps, des résultats obtenus». Ainsi, au-delà des résultats à court terme, la durabilité permet de prendre en compte le fait que «les enseignants et les autres acteurs espèrent sans doute produire des effets à long terme et inscrire la formation dans un horizon temporel d'une certaine ampleur: les formations scolaires ne visent pas uniquement à permettre de réussir les examens de l'année scolaire en cours, mais à préparer les adultes de demain» (Demeuse & Stauven, 2006, p. 205).

La flexibilité d'un programme correspond, quant à elle, à sa capacité d'adaptation et de changement. Cette qualité transversale est très utile et est envisagée comme une «méta» qualité. Elle prend en compte plusieurs données. Tout d'abord, il est question de déceler les améliorations potentielles de la politique prioritaire et ensuite, de mesurer la manière dont cette même politique serait capable de les mettre en œuvre, sans se dénaturer complètement.

Plusieurs constats peuvent être établis dans l'analyse de la flexibilité. La «flexibilité efficace» (Bouchard & Plante, 2002) correspond à la meilleure des situations car la politique évaluée dispose d'une capacité suffisante pour mettre en place les améliorations proposées. Par contre, on parle de «flexibilité en rupture» lorsque des lacunes sont constatées et qu'une solution ne paraît pas envisageable. C'est la flexibilité qui permet d'assurer la pérennité sans nécessiter de réformes en profondeur. Cela ne signifie pas que les programmes doivent être flous et les enseignants ou les formateurs versatiles, mais qu'un programme bien pensé, en fonction de l'identification de besoins, permet une certaine marge de

manœuvre à des acteurs réflexifs. Dit autrement, un programme doit permettre des ajustements nécessaires, sans un détricotage complet et une remise en cause fondamentale, en fonction de fluctuations contextuelles. Il doit, par exemple, permettre de tenir compte de réalités locales et de contraintes prévisibles sans qu'il soit nécessaire de procéder à une refonte de l'ensemble. Ainsi, l'arrivée de nouveaux groupes d'élèves ne doit pas nécessairement remettre en cause la totalité du programme.

Conclusion

Nous avons tenté d'attirer l'attention sur certains problèmes posés par l'évaluation des politiques d'éducation prioritaire, problèmes à propos du choix des méthodes à utiliser de manière à évaluer une politique, mais aussi du caractère multidimensionnel de la «qualité» d'un organisme, d'un programme ou d'une politique éducative. 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), par exemple lorsque des objectifs opérationnels et mesurables doivent être définis afin de pouvoir évaluer l'efficacité ou l'efficience d'une politique. 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 et que, au moins formellement les évaluations prolifèrent, même si elles ne présentent pas toujours toutes les garanties méthodologiques ou ne sont pas nécessairement adéquates par rapport aux objectifs. On ne peut, par exemple, pas évaluer l'efficacité d'une politique avant de l'avoir mise en place complètement... ce qui est sans doute un truisme... mais il est souvent ignoré par les commanditaires dont la durée de vie (politique) ne peut pas s'accommoder de la durée nécessaire à une évaluation sérieuse.

S'il ne fait aucun doute que les politiques éducatives devraient 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. L'expérimentation en éducation est une activité délicate, à la fois parce qu'elle pose des problèmes éthiques – ces problèmes se posent en médecine également –, 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.

Un autre argument à prendre en compte est la grande difficulté à évaluer

certaines qualités transversales, plus particulièrement celles concernant la question des résultats des PEP, et ce pour deux raisons principales (Demeuse, Friant & Nicaise, 2009). D'abord, les intentions et objectifs des PEP sont souvent déclinés en termes peu opérationnels, tel que le fait d'*assurer à tous les élèves des chances égales d'émancipation sociale* (objectif de la politique d'éducation prioritaire en Communauté française de Belgique): on comprend vite que l'évaluation du degré d'atteinte d'un tel objectif pose des problèmes insurmontables aux évaluateurs et qu'elle supposerait de définir des objectifs intermédiaires qui n'ont pas nécessairement été envisagés lors de la définition de cette politique. L'évaluation de l'efficacité, de l'efficience et de la durabilité des PEP se verrait ainsi facilitée par la déclinaison en objectifs plus précis et plus aisés à opérationnaliser. Ensuite, même dans le cas où les objectifs de la PEP sont opérationnalisés (par exemple, en termes de compétences des élèves), il est souvent impossible de disposer d'une mesure d'entrée, c'est-à-dire de données utilisables avant la mise en œuvre de la PEP, et parfois même d'une mesure de sortie, notamment lorsque les réformes succèdent rapidement aux réformes, ce qui hypothèque toute évaluation.

On le voit, si une réflexion méthodologique rigoureuse est possible, cette réflexion vient souvent trop tard dans un domaine où l'idéologie l'emporte sur la raison, alors même que beaucoup de politiques s'en défendent. Cela n'enlève rien à l'urgence d'agir que de mettre en place une méthodologie d'évaluation crédible, mais son absence affaiblit considérablement des actions finalement assez modestes face à des problèmes sociétaux très importants.

Notes

- 1 Voir par exemple Smith (1997) pour une discussion du terme «modèle», ou de celui de «théorie», dans le contexte de l'évaluation.
- 2 «Évaluation conduite par un évaluateur qui provient de l'extérieur de l'organisation où prend place l'étude» (Nadeau, 1988, p. 419).
- 3 «Évaluation conduite par un membre de l'organisation où l'étude prend place» (Nadeau, 1988, p. 419).
- 4 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.
- 5 Voir par exemple l'ouvrage, en langue française, de Nadeau (1988) à propos de l'évolution des modèles et des méthodes, principalement dans la littérature scientifique en langue anglaise.
- 6 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).
- 7 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.
 - 8 Pour une illustration des débats autour du programme *Success for All* de Slavin, Madden, Dolan et Wasik (1996), on peut consulter l'article de Demeuse et Denooz (2001).
 - 9 Frandji (2008) met ainsi en garde contre, d'une part, l'approche «cavalier solitaire» (Osborn, 2007) ou «fausse comparaison», juxtaposant des analyses réalisées séparément par des équipes différentes, sans tentative de synthèse, et, d'autre part, l'approche «safari» (Osborn, 2007), où l'étude est réalisée par une seule équipe nationale.
 - 10 Nous nous référons ici à un avantage de l'approche comparative dans le cadre de l'évaluation des politiques d'éducation. Dans une visée plus large, l'approche comparative telle qu'adoptée par l'équipe EuroPEP permet de mettre en œuvre, à propos des politiques prioritaires, une «description de leur fonctionnement, mais aussi, dans la perspective d'une sociologie de la connaissance, une analyse de la manière dont elles se construisent, se définissent et sont pensées» (Frandji & Rochex, 2011, p. 16).
 - 11 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)».
 - 12 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.
 - 13 «Si l'on veut jouer avec les mots, les acronymes (clair – éclair) faciliteraient ici la tâche, comme cela s'observe d'ailleurs sur le terrain: ses finalités questionnent, voire inquiètent, ses critères de mise en œuvre sont pour une part opaques; peu de cadres de l'éducation nationale acceptent de le discuter officiellement; des personnels d'établissements qui manifestent leur inquiétude doivent semble-t-il aller jusqu'à la grève pour se faire entendre; il semble par contre bel et bien se développer à grande vitesse, mais avant même le résultat d'une éventuelle évaluation que l'on aurait pu croire accompagner cette première étape d'expérimentation» (Frandji, 2011, p. 1).

Références bibliographiques

- Antoniou, L., Dyson, A. & Raffo, C. (2008). Entre incantation et fébrilité: Les nouvelles politiques d'éducation prioritaire en Angleterre (1997-2007). In M. Demeuse, D. Frandji, D. Greger & J.-Y. Rochex (Éd.), *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en oeuvre, débats* (pp. 37-84). Lyon: INRP.
- Aubert-Lotarski, A., Demeuse, M., Derobertmeasure, 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'Éducation*, 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* (pp. 124-131). Paris: L'Harmattan.
- Bollens, J., Van de Velde, V., Cnudde, V., Vanobbergen, B., Vansieleghem, N., Douterlun-

- gne, M., Nicaise, I. & Verhaeghe, J.-P. (1998). *Zorgverbreding in het basis- en secundair onderwijs: Een zoektocht naar financieringscriteria*. Leuven: HIVA / RUG, Vakgroep Onderwijskunde
- Borman, G. D., Stringfield, S. C. & Slavin, R. E. (Éd.). (2001). *Title I. Compensatory education at the crossroads*. Mahwah (NJ): Lawrence Erlbaum Associates.
- Bouchard, C. & Plante, J. (2002). La qualité: Mieux la définir pour mieux la mesurer. *Les Cahiers du Service de Pédagogie expérimentale*, 11-12, 219-236.
- Bouchat, T.-M., Delvaux, B. & Hindryckx, G. (2005). *Discrimination positive et mobilité scolaire* (Rapport de recherche). Cerisis-Université catholique de Louvain et SPE-Université de Liège.
- Bouchat, T.-M., Delvaux, B. & Hindryckx, G. (2008). Mobilité scolaire et composition sociale des écoles. Le cas de l'enseignement fondamental en Communauté française de Belgique. *Les cahiers de recherche en éducation et formation*, 62, 1-41.
- Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles: De Boeck Université.
- Bronfenbrenner, U. (1979). *The ecology of human development. Experiments by nature and design*. Cambridge: Harvard University Press.
- Crahay, M. (2005). Pourquoi l'expérimentation en éducation? In M. Demeuse, A. Baye, M. H. Straeten, J. Nicaise & A. Matoul (Éd.), *Vers une école juste et efficace. 26 contributions sur les systèmes d'enseignement et de formation* (pp. 95-114). Bruxelles: De Boeck Université.
- Dauvisis, M. C. (2006). L'instrumentation en évaluation. *Mesure et évaluation en éducation*, 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. & Monseur, C. (1999). Analyse critique des indicateurs déterminant l'attribution des moyens destinés à la politique de discrimination positive en Communauté française de Belgique. *Mesure et évaluation en éducation*, 22, (2-3), 97-127.
- Demeuse, M. & Strauven, C. (2006). *Développer un curriculum d'enseignement ou de formation. Des options politiques au pilotage*. Bruxelles: De Boeck Université.
- Demeuse, M., Demierbe, C. & Friant, N. (2011). Quelle évaluation des politiques d'éducation prioritaire en Europe? Un cadre d'analyse et son application. In M. Demeuse, D. Frandji, D. Greger & J.-Y. Rochex (Éd.), *Comparaison des politiques d'Éducation prioritaire en Europe. Quel devenir pour l'égalité scolaire?* (pp. 293-316). Lyon: ENS.
- Demeuse, M., Derobertmeasure, A. & Friant, N. (2010). Differentiated financing of schools in French-speaking Belgium: Perspectives for regulating a school quasi-market. *Educational Research and Evaluation*, 16, (2), 197-216.
- Demeuse, M., Friant, N. & Nicaise, I. (2009). Les politiques d'éducation prioritaire: Donner plus à ceux qui ont moins. In I. Nicaise, E. Desmedt & M. Demeuse (Éd.), *Une école réellement juste pour tous! Seize pistes de réforme et d'action* (pp. 387-419). Waterloo: Plantyn.
- Demeuse, M., Crepin, F., Jehin, M. & Matoul, A. (2006). Behind the positive discrimination in French Community of Belgium: Central criteria vs. local actions. In L. Moreno Herrera, G. Jones & J. Rantala (Éd.), *Enacting equity in education: Towards a comparison of equitable practices in different European local contexts* (pp. 59-79). Research Centre for Social Studies Education, University of Helsinki.
- Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (Éd.). (2008). *Les politiques d'éducation prioritaire en Europe: Conceptions, mise en œuvre, débats*. Lyon: Institut national de Recherche pédagogique.
- Demeuse, M., Frandji, D., Greger, D. & Rochex, J.-Y. (Éd.). (2011). *Comparaison des politiques d'Éducation prioritaire en Europe. Quel devenir pour l'égalité scolaire?* Lyon: ENS.
- Demeuse, M., Matoul, A., Schillings, P. & Denooz, R. (2005). De quelle efficacité parle-

- t-on? In M. Demeuse, A. Baye, M. H. Straeten, J. Nicaise, A. Matoul (Éd.), *Vers une école juste et efficace. 26 contributions sur les systèmes d'enseignement et de formation* (pp. 15-27). Bruxelles: De Boeck Université.
- Epstein, E. (1988). The problematic meaning of 'Comparison' in Comparative Education. In J. Schriewer & B. Holmes (Éd.), *Theories and methods in Comparative Education*. Frankfurt: Peter Lang.
- Frاندji, D. (2008). Pour une comparaison des politiques d'éducation prioritaire en Europe. In M. Demeuse, D. Frاندji, D. Greger & J.-Y. Rochex (Éd.), *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en œuvre, débats* (pp. 9-34). Lyon: Institut national de Recherche pédagogique.
- Frاندji, D. (2011). Les points obscurs du programme Clair: Une nouvelle politique scolaire? Contribution au colloque du SNES: *Redonner un avenir au collège*. Consulté le 5 mai 2011 dans http://www.snes.edu/IMG/pdf/Programme_clair_et_devenir_de_l_education_prioritaire.pdf
- Frاندji, D. & Rochex, J.-Y. (2011). Du devenir des politiques d'éducation prioritaire à la recherche de possibles, introduction de l'ouvrage. In M. Demeuse, D. Frاندji, D. Greger & J.-Y. Rochex (Éd.), *Les politiques d'éducation prioritaire en Europe. Tome II. Quel devenir pour l'égalité scolaire?* (pp. 11-51). Lyon: ENS.
- Gazibo, M. & Jenson, J. (2004). *La politique comparée. Fondements, enjeux et approches théoriques*. Montréal: Presses de l'Université.
- Hall, P. A. & Taylor, R. (1997). La science politique et les trois néo-institutionnalismes. *Revue française de science politique*, 47, (3-4), 469-496.
- Legendre, R. (2005). Dictionnaire actuel de l'éducation (3e éd.). Montréal (Québec): Guérin.
- Léon, A., Cambon, J., Lumbroso, M. & Winnykamen, F. (1977). *Manuel de psychopédagogie expérimentale*. Paris: Presses Universitaires de France.
- Maratou-Alibranti, L., Teperoglou, A. & Tsiganou, I. (2006). *To elliniko sholio me tin avgi tou 21ou eona. Ekpedefikis anaggis, provlimata ke prooptikes* [L'école grecque à l'aube du 21e siècle. Besoins éducatifs, problèmes et perspectives]. Athènes: Gutenberg.
- Meuret, D. (1994). L'efficacité de la politique des zones d'éducation prioritaire. *Revue française de pédagogie*, 109, 41-64
- Mingat, A. (1999). Évaluation des politiques publiques en éducation: Quelques balises pour leur conduite. In G. Solaux (Éd.), *L'évaluation des politiques d'éducation* (pp. 33-47). Dijon: CRDP de Bourgogne.
- 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. (2007b). *Les nouvelles politiques éducatives. La France fait-elle les bons choix?* Paris: Presses Universitaires de France.
- Nadeau, M. A. (1988). *L'évaluation de programme. Théorie et pratique* (2e éd.). Québec: Les Presses de l'Université Laval.
- Nisbet, J. (1997). Policy-oriented Research. In J. P. Keeves (Éd.), *Educational research methodology, and measurement: An international handbook* (pp. 211-217). Cambridge: Pergamon.
- OFSTED (Office for standards in education). (2003). *Education action zones: Tackling difficult issues in round 2 zones*. Rapport de recherche publié par l'Office for standards in Education (référence HMI 1711). Consulté le 5 mai 2011 dans <http://www.ofsted.gov.uk/resources/education-action-zones-tackling-difficult-issues-round-2-zones>
- Osborn, M. (2007). Promouvoir la qualité: Comparaisons internationales et questions méthodologiques. *Education et sociétés*, 18, 163-180.
- Power, S., Whitty, G., Gewirtz, S., Halpin, D. & Dickson, M. (2004). Paving a 'third way'? A policy trajectory analysis of Education Action Zones. *Research Papers in Education*, 19, (4), 453-476.
- Rochex, J.-Y. (2008). Les politiques d'éducation prioritaire en Europe, d'un «âge» et d'un

- pays l'autre. In M. Demeuse, D. Frandji, D. Greger & J.-Y. Rochex (Éd.), *Les politiques d'éducation prioritaire en Europe: Conceptions, mises en œuvre, débats* (pp. 409-451). Lyon: Institut national de recherche pédagogique.
- 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 (Éd.), *Educational research methodology, and measurement: An international handbook* (pp. 217-224). Cambridge: Pergamon.
- Solaux, G. (1999). Conclusion. In G. Solaux (Éd.), *L'évaluation des politiques d'éducation* (pp. 179-187). Dijon: CRDP de Bourgogne.
- Sowell, T. (2004). *Affirmative action around the world. An empirical study*. 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.

Mots-clés: Politiques éducatives, évaluation, éducation prioritaire, comparaisons internationales, systèmes éducatifs européens

Die Bewertung der prioritären Bildungspolitik in Europa: eine methodische Herausforderung

Zusammenfassung

Die Bewertung der Bildungspolitik ist keine einfache Sache, einerseits weil sie auf methodische Schwierigkeiten stößt, andererseits weil sie gewisse Projekte, die maßgeblich auf einer ideologischen Basis aufgebaut sind – auch wenn dies von den Erfindern zurückgewiesen wird – bedroht oder zu bedrohen scheint. Es ist aber notwendig diese politischen Prozesse bewerten zu können, besonders wenn sie benachteiligte Bevölkerungsgruppen betreffen. Dieser Artikel, der auf eine Untersuchung aufbaut, die in 8 europäischen Ländern durchgeführt wurde (EuroPEP), illustriert die beobachteten Probleme, zeigt aber auch mögliche Lösungsvorschläge auf. Er verdeutlicht die Notwendigkeit einer globalen methodischen Betrachtung und hinterfragt, unter anderem, die Art und Weise der Definition einer «Best Practice» im Bereich der Bildungspolitik.

Schlagworte: Bildungspolitik, Bewertung, Prioritäre Bildung, internationale Vergleiche, europäische Bildungssysteme

