

Les méthodes des enquêtes internationales et leurs fonctions politiques.

L'exemple de la France face à PISA (1995-2008)

Xavier Pons

Volume 34, numéro 2, 2011

Date de réception : 11 juin 2011

Date de réception de la version finale : 5 septembre 2011

Date d'acceptation : 8 septembre 2011

URI : <https://id.erudit.org/iderudit/1024849ar>

DOI : <https://doi.org/10.7202/1024849ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Pons, X. (2011). Les méthodes des enquêtes internationales et leurs fonctions politiques. L'exemple de la France face à PISA (1995-2008). *Mesure et évaluation en éducation*, 34(2), 57–85. <https://doi.org/10.7202/1024849ar>

Résumé de l'article

Le choix de la méthode à l'origine d'une enquête internationale n'est pas qu'une question d'ordre technique ou scientifique mais a des implications politiques majeures. Dans une perspective de sociologie qualitative de l'action publique, cet article analyse les fonctions politiques remplies par la méthode au fondement du programme PISA de l'OCDE entre 1995 et 2008 à travers l'expérience particulière de la France. L'article met en évidence que les fonctions politiques remplies par la méthode diffèrent selon les étapes de l'instrumentation de l'action publique. Lors de la conception de l'enquête, la méthode constitue un outil puissant de problématisation de l'action publique. Lors de sa mise en oeuvre, elle remplit une fonction de contrôle social. La réception de PISA en France enfin montre que la méthode a un pouvoir très important de structuration du débat public dans la mesure où elle prédétermine fortement les usages politiques et cognitifs qui peuvent en être faits.

Les méthodes des enquêtes internationales et leurs fonctions politiques. L'exemple de la France face à PISA (1995-2008)

Xavier Pons

Université Paris-Est Créteil

MOTS CLÉS: PISA, méthode des comparaisons internationales, politiques éducatives, instruments d'action publique, débat public

Le choix de la méthode à l'origine d'une enquête internationale n'est pas qu'une question d'ordre technique ou scientifique mais a des implications politiques majeures. Dans une perspective de sociologie qualitative de l'action publique, cet article analyse les fonctions politiques remplies par la méthode au fondement du programme PISA de l'OCDE entre 1995 et 2008 à travers l'expérience particulière de la France. L'article met en évidence que les fonctions politiques remplies par la méthode diffèrent selon les étapes de l'instrumentation de l'action publique. Lors de la conception de l'enquête, la méthode constitue un outil puissant de problématisation de l'action publique. Lors de sa mise en œuvre, elle remplit une fonction de contrôle social. La réception de PISA en France enfin montre que la méthode a un pouvoir très important de structuration du débat public dans la mesure où elle prédétermine fortement les usages politiques et cognitifs qui peuvent en être faits.

KEY WORDS: PISA, methods of international surveys, education policies, regulation tools, public debate

Choosing the method at the basis of an international survey is not only a technical or a scientific issue since the method has deep political implications. This article analyses the political functions carried out by the method at the origin of the OECD's PISA program through the particular national experience of France (1995-2008). Adopting the standpoint of a qualitative sociology of education policies, it shows that the political functions of the method are different according to the sequences of the policy process. During the conception of PISA at the international level, the method strongly contributes to formulate the problems that must be addressed, how and by who. During the implementation in France, the method is a powerful tool of social control. Concerning the reception of PISA in France, the method highly structures the national policy debate by pre-determining the possible cognitive and political uses of the survey.

PALAVRAS-CHAVE: PISA, métodos de comparação internacionais, políticas educativas, instrumentos de acção pública, debate público

A seleção de um método para a realização de um inquérito internacional não é apenas uma questão de ordem técnica ou científica, é também uma questão com implicações políticas significativas. Numa perspetiva de sociologia qualitativa da ação pública, este artigo analisa as funções políticas do método que está na base do programa PISA da OCDE entre 1995 e 2008 através da experiência particular da França. O artigo evidencia que as funções políticas do método diferem segundo as etapas da instrumentação da ação pública. Durante a conceção do inquérito, o método constitui um instrumento que contribui intensamente para a problematização da ação pública. Durante a implementação, o método realiza uma função de controlo social. A receção do PISA em França mostra, no final de contas, que o método tem um poder muito importante na estruturação do debate público, na medida em que pré-determina fortemente os seus usos políticos e cognitivos.

Note de l'auteur – Toute correspondance peut être adressée comme suit : Xavier Pons, Largotec, UPEC, Bâtiment T, 1^{er} étage, bureau 112, Centre multidisciplinaire de Créteil, 61, avenue du Général de Gaulle, 94000 Créteil, France, téléphone : +33(0)6 28 78 78 16, ou par courriel à l'adresse suivante : [ponsx@wanadoo.fr].

Introduction

Les comparaisons internationales des acquis des élèves produisent des connaissances spécifiques qui donnent lieu à des publications et des études complémentaires et qui s'insèrent plus ou moins bien dans les traditions académiques qui lui préexistent comme l'éducation comparée (Mons, 2007a; Novoá, 2001). Mais elles sont aussi un nouvel instrument d'action publique emblématique des changements de modes de régulation qui affectent la plupart des systèmes éducatifs de l'OCDE (Grek, 2009; Henry, Lingard, Rizvi, & Taylor, 2000; Mons, 2008; Morgan, 2007), contribuant ainsi à institutionnaliser sur les plans mondial et supranational (européen, par exemple) une gouvernance douce de l'éducation (Lawn, 2006) et une gouvernance par les résultats (Harris & Herrington, 2006; Maroy, 2006). Très peu de travaux francophones tirent les conséquences de cette dualité en envisageant les enquêtes comparatives internationales non plus seulement comme des opérations de connaissance mais aussi comme des opérations politiques, et plus précisément comme des modes particuliers d'instrumentation de l'action publique (Lascoumes & Le Galès, 2004). Il peut s'agir de l'action publique constituée par la préparation et l'administration de l'enquête internationale elle-même, mais aussi des actions publiques qui se nouent autour des débats suscités par ses résultats et des mesures prises en référence à son contenu.

Certes, plusieurs critiques d'ordre méthodologique ont pu être adressées aux comparaisons internationales des acquis des élèves par les chercheurs et les statisticiens en éducation depuis la fin des années 1980 : manque de fondement théorique et méthodologique (notamment en matière de choix des pays participants), mesure de certaines dimensions seulement des disciplines scolaires, agrégation problématique de scores dans une même discipline à l'échelle d'un pays, échantillonnages différents d'une enquête à l'autre entravant la comparabilité des résultats dans le temps, persistance d'importants problèmes de traduction des items d'une langue à l'autre, etc.¹ Mais ces critiques concernent le plus souvent la pertinence logique des choix méthodologiques eux-mêmes et non leurs significations et implications politiques. Certains chercheurs évoquent ces dernières, en dénonçant par exemple le caractère impérialiste des comparaisons internationales qui imposeraient un modèle unique et uniforme d'éducation (Mollis & Marginson, 2002), leur focalisation excessive sur l'État-nation dans un contexte de

globalisation (Crossley, 1999) ou leur prise en compte insuffisante des contextes nationaux empêchant ainsi tout transfert de politiques ou de bonnes pratiques éducatives d'un contexte éducatif à un autre (Goldstein, 1995). Mais, le plus souvent déduites d'une critique méthodologique interne, ces dénonciations reposent rarement sur une enquête empirique auprès des évaluateurs, chercheurs et experts participant à ces enquêtes. La sociologie des sciences, notamment celle des réseaux socio-techniques et de la traduction (Callon, 1986, 1989; Callon & Latour, 1985), qui pourrait favoriser ce type d'analyse, a donné lieu à peu d'applications dans le champ de la recherche francophone en éducation, encore moins en ce qui concerne les enquêtes internationales (voir par exemple Normand, 2010). Quant aux spécialistes francophones de sociologie de l'action publique et de l'analyse des politiques publiques qui ont abordé le thème de l'évaluation, ils se sont principalement intéressés à son degré d'institutionnalisation et aux formes prises par cette dernière (Duran, 2002; Duran & Monnier, 1992; Jacob & Varone, 2004), à ses effets sur le fonctionnement des administrations publiques (CURAPP, 1993) et à ses liens avec les expériences passées de rationalisation de l'action administrative (Spenehauer, 1998), le tout pour interroger les effets de la nouvelle gestion publique et l'évolution du rôle des États (Bézès, 2009), de sorte qu'à quelques exceptions près (Lascoumes & Setbon, 1996), les méthodes d'évaluation elles-mêmes sont peu remises en question.

Objet, méthodologie et limites

L'objet de cet article est de contribuer à cette interrogation en analysant, dans une perspective de sociologie de l'action publique, les fonctions politiques remplies de fait par la méthode de l'enquête PISA de l'OCDE² lors de sa mise en œuvre par la France entre 1995 et 2008, c'est-à-dire depuis la phase de préparation de PISA sur un plan international dans le cadre du réseau A du projet INES jusqu'à la dernière année couverte par nos travaux de recherche. Par méthode, nous désignons l'ensemble des choix théoriques et pratiques réalisés par un groupe d'évaluateurs au moment d'appréhender son objet d'analyse, qu'il s'agisse de l'organisation concrète de la démarche d'évaluation, d'une problématisation plus ou moins théorique ou de l'ensemble des techniques d'enquête concrètement utilisées (Pons, 2010).

Le présent article repose sur des matériaux rassemblés dans le cadre de deux recherches complémentaires au moyen de méthodes qualitatives (état de la documentation scientifique et professionnelle, entretiens semi-directifs,

analyse systématique de publications officielles et construction d'une statistique descriptive). La première est une thèse de doctorat, effectuée entre 2004 et 2008, portant sur l'évolution de la politique française d'évaluation (Pons, 2008). Dans cette recherche, nous avons pu analyser l'activité des membres de la DEP (32 entretiens sur les 98 menés), notamment ceux qui représentent la France au sein des divers programmes de l'OCDE, et interroger des membres de cette organisation internationale. La deuxième s'inscrit dans le cadre du projet européen *Know&Pol* (2006-2011) coordonné par l'Université catholique de Louvain et visant à analyser le rôle de la connaissance dans la régulation des systèmes d'éducation et de santé dans divers pays européens. L'une des orientations de ce projet était consacrée à la conception de PISA sur le plan international et à sa réception dans six pays européens. Cette recherche a été l'occasion d'évaluer la contribution des représentants de la France au sein du programme PISA (Pons, 2009) et d'analyser avec Nathalie Mons, sur la base d'un corpus multiple (trentaine d'entretiens, exploitation de près de 300 dépêches d'une agence de presse spécialisée, analyse d'un corpus de 120 titres de presse nationale et régionale), la réception de cette enquête internationale dans le débat public sur le plan national (Mons & Pons, 2010).

Nous faisons l'hypothèse que les fonctions politiques remplies par la méthode diffèrent selon les étapes de l'instrumentation de l'action publique car chaque étape met en jeu un système d'action spécifique (avec ses acteurs, ses institutions, ses intérêts, etc.) au sein duquel l'ensemble des problèmes politiques posés par le choix et l'usage de la méthode peuvent différer. C'est pourquoi nous distinguons trois séquences d'action publique classiques (conception, mise en œuvre et réception) qui, dans les faits, se recoupent et qui mettent chacune d'elle en évidence une fonction politique différente remplie par la méthode : une fonction de problématisation (conception), de contrôle social (mise en œuvre) et de structuration du débat public (réception). Une telle approche présente trois écueils potentiels qu'il convient au préalable d'évoquer afin de les neutraliser, d'explicitier notre point de vue (Weber, 1965) et de justifier nos choix méthodologiques.

Le premier écueil renvoie au *degré de technicité* de notre objet. Autant l'écrire clairement, nous ne sommes ni évaluateur, ni chercheur spécialiste des techniques statistiques, psychométriques ou même didactiques à l'origine des comparaisons internationales des acquis des élèves. Si ce fait peut constituer un obstacle à la compréhension fine et complète des enjeux liés au

choix d'une méthode, il est aussi une occasion de proposer un regard externe complémentaire et relativement décalé par rapport au point de vue privilégié par les spécialistes du champ. Aussi, la focalisation sur l'enquête PISA ne s'explique-t-elle pas par un quelconque jugement technique ou scientifique sur sa qualité méthodologique mais par les nombreux enjeux politiques qui lui sont de fait associés. Cette enquête internationale est explicitement conçue et présentée comme un outil d'aide à la décision, supposée en particulier jouer un rôle de catalyseur (Carvalho & Costa, 2009). Elle s'inscrit dans une stratégie spécifique mise en avant par une organisation internationale, l'OCDE, qui, sans avoir initialement de mandat officiel en matière d'éducation (Papadopoulos, 1994), a progressivement fait de la comparaison, de la mise en évidence des bonnes pratiques et de l'analyse entre pairs (*peer review*), des vecteurs majeurs de diffusion de nouvelles politiques éducatives (Mons, 2007b). Enfin, l'enquête PISA a connu une percée médiatique importante à partir de 2005 en France, qui a invité un nombre croissant d'acteurs du débat public à revoir les comparaisons internationales et qui a participé ainsi à une réorientation de la politique française d'évaluation (Pons, 2010).

Le deuxième écueil est celui de *l'ethnocentrisme* dans la mesure où cet article, rédigé par un sociologue français, n'évoque que le cas de la participation de la France. Ce choix découle de considérations méthodologiques précises qui renvoient aux matériaux accumulés lors de précédentes recherches centrées sur ce pays et aux particularités de ce dernier en matière de coopération internationale. La France est connue pour ses revendications en matière d'exception culturelle dans le domaine des arts, de la culture ou de l'éducation. Elle entre dans le programme PISA après le scandale suscité en mai 1995 par son retrait de l'enquête IALS³ après avoir commandité plusieurs contre-expertises au motif que la méthode n'est pas apparue suffisamment solide aux yeux de ses représentants, les membres de la Direction de l'évaluation et de la prospective (DEP⁴) du ministère de l'Éducation nationale (Pons, 2008). Enfin, dans le prolongement de la présidence française de l'Union européenne, la DEP a pris la tête, en novembre 1996, d'un réseau d'évaluateurs européens avec lesquels elle a conçu une méthode alternative d'évaluation internationale visant non pas à administrer les mêmes exercices traduits dans des langues différentes mais à construire des critères de comparabilité d'exercices conçus dans les langues maternelles. Cette méthode, présentée au comité chargé d'expertiser les réponses à l'appel d'offre de PISA (*PEACE project*), a été classée dernière (Bottani & Vrignaud, 2005). Pour ces

trois raisons, la France constitue un terrain d'investigation privilégié des oppositions d'ordre politique suscitées par la méthode retenue dans PISA, un révélateur des fonctions politiques remplies par celle-ci.

Le troisième écueil, lié aux deux précédents, est celui de *l'implicite normatif* qui, en donnant principalement la parole aux acteurs français, laisserait entendre ce qu'aurait dû être la « vraie » méthode à l'origine de PISA. Or, les données « sont les discours évoquant les choses et non les choses évoquées par les discours » (Schwartz, 1993, p. 282). Les faits mentionnés dans les propos de nos interlocuteurs français ne sont donc pas repris à notre compte mais considérés comme des éléments révélant les luttes, tensions et controverses propres à toute action politique.

Résultats

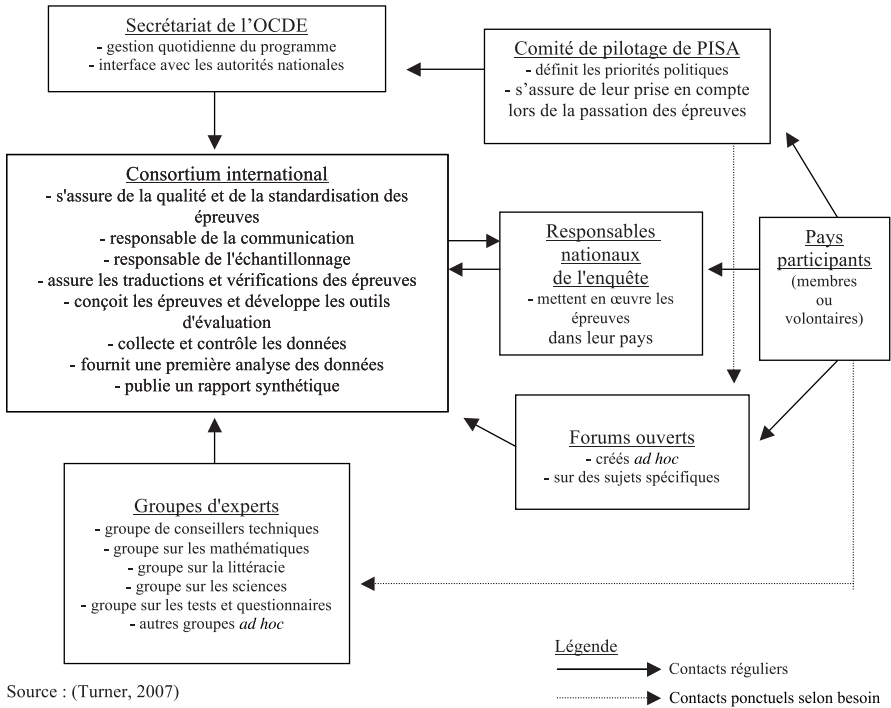
La méthode comme problématisation de l'action publique

Si la méthode au fondement de l'enquête PISA est constamment ajustée au fur et à mesure de sa mise en œuvre, les principaux choix méthodologiques sont effectués au cours de la phase de conception, notamment entre 1995 et 1999. Cette phase n'est pas qu'une étape technique relativement évidente, elle est aussi l'occasion de définir un certain nombre de règles de fonctionnement qui ne sont pas neutres d'un point de vue politique. Pour reprendre les termes de Michel Callon (1986, p. 180-185), la méthode est un élément décisif de « problématisation » de l'action publique. Elle formule de manière particulière les problèmes à résoudre, elle définit les « points de passage obligés » de la réflexion collective comme de la mise en œuvre institutionnelle de l'enquête et elle contribue à sélectionner de fait les interlocuteurs et les thèmes de discussion scientifique politiquement légitimes (la méthode comme « entre-définition des acteurs »).

PISA : principales caractéristiques méthodologiques

La conférence de Budapest de mai 1997 permet, après deux ans de réflexion au sein du réseau A du projet INES dont les conclusions sont mentionnées dans le document DEELSA/ED/CERI/CD(97)4 du 28 mars 1997 (Bottani & Vignaud, 2005), d'arrêter les grands principes de l'enquête, qu'ils soient d'ordre organisationnel (voir le schéma 1) ou d'ordre méthodologique : focalisation sur la littéracie, interrogation des compétences des élèves des pays de l'OCDE (et de pays partenaires volontaires) à l'âge de 15 ans qui

correspond à la fin de la scolarité obligatoire dans un grand nombre de pays, cycle d'interrogations de trois ans avec à chaque session une discipline dominante à laquelle sont consacrés les deux tiers des items, etc.



Source : (Turner, 2007)

Schéma 1. *L'architecture institutionnelle de PISA*

L'appel d'offres international lancé le 9 octobre 1997 et remporté par ACER, un centre de recherche indépendant australien créé en 1930, constitue une étape clé en termes de méthodologie. Si l'enquête repose sur certaines techniques relativement classiques en psychométrie et en statistique (technique des cahiers tournants, modèle de réponse à l'item, échantillonnage aléatoire par degré et stratifié, possible administration de questionnaires contextuels aux élèves, aux parents et aux chefs d'établissement pour envisager un certain nombre de variables explicatives), retenir la proposition d'ACER revient de fait à privilégier un modèle de réponse à l'item particulier, dérivé du modèle de Rasch et intégré dans un logiciel proposé par cet organisme (le logiciel *ConQuest*) qui repose sur trois hypothèses majeures qui sont autant de

conditions de validité du test (Vrignaud, 2006) : l'unidimensionnalité (la réussite à un item ne dépend que du niveau de compétence de l'individu), l'indépendance conditionnelle des items (la réponse d'un individu à un item ne dépend pas de ses réponses aux autres items de l'épreuve, et ce, lorsque le niveau de compétence est contrôlé) et l'égal pouvoir discriminant des items (quels que soient les items passés par un sujet, on obtiendra une même estimation de sa compétence).

Plusieurs réunions sont ensuite organisées en 1998 et 1999 associant le consortium international dirigé par ACER, le secrétariat de l'OCDE, le comité directeur de PISA, les groupes d'experts conseillers et parfois le comité directeur du CERI pour concevoir le protocole d'enquête et les items selon une répartition formelle des tâches présentée dans le schéma 1. Tous les pays sont invités à proposer des items en plus de ceux prévus par le consortium. Pour PISA 2000 par exemple, il y en avait en tout plus de 800, mais seulement 372 items, remaniés après une première sélection de 586 items, sont retenus (OCDE, 2002). Les experts nationaux doivent ensuite apprécier la pertinence de ces 372 items au regard des particularités culturelles de leur pays. Une banque de données recense tous les items, ainsi que les commentaires effectués pour chacun d'eux par tous les pays participants. L'expérimentation des items (un an avant la passation des épreuves) permet en outre d'améliorer «*l'équivalence culturelle*» des items (OCDE, 2002, p. 21). Des consignes strictes de traduction sont données. Les items soupçonnés de comporter un biais culturel (ceux qui présentent un «*fonctionnement différentiel*») sont «*revus en concertation avec les experts nationaux des pays concernés*» (OCDE, 2002, p. 23). Le choix final des items est arrêté à l'unanimité par le comité directeur, sur la base des propositions faites par le consortium international. Le protocole est testé en 1999 et mis en œuvre pour la première fois l'année suivante.

Partialités politiques de la méthode

Ces choix et caractéristiques méthodologiques ont plusieurs implications d'ordre politique. La première est que le modèle de réponse à l'item retenu par le consortium international repose sur trois hypothèses fortes qui ont en commun de donner un pouvoir décisif à ceux qui conçoivent les items puisque, une fois ces derniers choisis, ils vont être pensés, en l'état de la méthodologie retenue lors du premier cycle PISA, comme étant non biaisés sur le plan culturel. L'hypothèse d'indépendance conditionnelle des items signifie que la réponse d'un individu à un item ne dépend pas de ses réponses aux autres

items, et ce, lorsque le niveau de compétence est contrôlé. Il n'y aurait donc aucun effet de surprise ou, à l'inverse, de réconfort provoqué par la teneur des items, ni même aucun effet psychologique d'entraînement. Un élève rassuré par des items proches de ce qu'on lui enseigne d'habitude ne serait donc pas en meilleure disposition psychologique pour répondre aux items suivants qu'un élève qui découvre leur logique et leurs implicites. D'après l'hypothèse d'unidimensionnalité, la réussite à un item ne dépend pas du contexte ou des caractéristiques sociales du groupe auquel appartient l'élève (Vrignaud, 2006). D'après l'hypothèse d'égal pouvoir discriminant des items, on obtient une même estimation de la difficulté d'un item quels que soient les groupes de sujets auxquels l'item a été administré. Le modèle est donc construit de telle manière que les conditions nécessaires à l'existence d'une situation permettant de conclure que la performance d'un élève à un item ne dépend que de sa compétence propre – ce qui, rappelons-le, est la raison principale à l'origine du choix des modèles de réponse à l'item pour ce type d'enquête – sont intégrées directement dans les hypothèses de départ et non testées *a posteriori*. Vrignaud (2006) regrette ainsi qu'un ensemble de travaux psychométriques ayant mis en évidence les différentes procédures (une vingtaine selon lui) nécessaires pour s'assurer du respect de ces hypothèses ne soient pas intégrés dans le protocole d'enquête de PISA. En termes politiques, cela signifie que les acteurs susceptibles de proposer des items recevables ont un grand pouvoir puisqu'en l'état de la méthodologie PISA lors du premier cycle d'enquêtes, s'ils réussissent à faire accepter un item, il est fort probable que celui-ci soit considéré comme universellement légitime.

Or, tous les acteurs (pays ou organismes) ne sont pas en égale capacité de proposer des items. Le tableau 1 qui donne la liste des contributeurs d'items pour PISA 2000 montre ainsi l'influence évidente des membres du consortium ACER, CITO et ETS (via les chiffres pour l'enquête IALS) qui proposent la majorité des items retenus au premier tour (333 sur 586). Les pays de l'Union européenne contribuent peu en comparaison d'autres pays ou continents. La France est dans une position intermédiaire. Contributeur faible au regard du nombre d'items (32), elle est le cinquième contributeur et fait partie des acteurs proposant des items dans les trois domaines évalués.

Tableau 1
Les contributeurs d'items pour PISA 2000

<i>Domaine / Pays</i>	<i>Lecture</i>	<i>Mathématiques</i>	<i>Sciences</i>	<i>Total</i>
ACER	123	3	29	155
Allemagne	2			2
Australie			4	4
Autriche	7	5		12
Belgique	13			13
Corée	6		8	14
Danemark	4			4
États-Unis	20	10	15	45
Fédération de Russie			4	4
Finlande	21			21
France	14	7	11	32
Grèce	6	3		9
Groupe CITO	51	34	20	130
Groupe République tchèque		2		2
IALS	48			48
Irlande	7			7
Italie	8			8
Japon	8			8
Norvège	7		6	13
Nouvelle-Zélande	13			13
Royaume-Uni	6			6
Suède	8	6	6	20
Suisse	6			6
TIMSS		5	5	10
Total	376	77	108	586

Source : (OCDE, 2002, p. 196)

Cette situation s'explique notamment – et c'est la deuxième implication politique que nous voulons souligner – par le fait que les choix méthodologiques effectués dans le cadre de PISA consacrent comme dominante sur le plan scientifique et légitime sur le plan politique une forme de connaissance particulière que nous proposons d'appeler *l'éducation comparée quantitative* (Pons, 2008) et que les pays participants sont plus ou moins en capacité de produire selon leurs traditions académiques, leur niveau d'expérience en matière d'évaluation des acquis des élèves, les conjonctures politiques à l'œuvre lors du lancement de l'enquête et l'état des rapports de force domestiques entre institutions et groupes professionnels⁵. En France, par exemple, la pratique de la psychométrie dans le secteur de l'éducation est à la fois dispersée et résiduelle malgré le dynamisme de quelques équipes

universitaires ou la réputation de certains organismes en matière de docimologie comme l'Inetop (Vrignaud, 2005). Les raisons sont aussi bien d'ordre académique (faible développement général des méthodes d'analyse quantitative dans la recherche en éducation, focalisation sur d'autres problématiques comme le fonctionnement des établissements scolaires, évolution de la psychologie différentielle qui se concentre progressivement sur les processus cognitifs et de moins en moins sur les nouveaux tests, transmission marginale des acquis de la psychométrie dans le cadre de cours optionnels dans les écoles nationales de statisticiens), que professionnel (privilège accordé par les conseillers d'orientation psychologues à une vision éducative de l'orientation des élèves pensés comme des acteurs de leur destinée scolaire), institutionnel (centralisation en France de la production de données par une direction d'administration centrale dont les membres sont inégalement formés à la psychométrie, suivi des enquêtes internationales par quatre organismes différents depuis les années 1960) et politique, le lancement de l'enquête PISA correspondant par exemple en France au ministère de Claude Allègre (1997-2000) qui s'est traduite par une mise entre parenthèses des travaux de la DEP (Pons, 2010).

Dans ce contexte, il reste difficile pour la France de pénétrer les organes de conception du programme PISA et de peser sur les choix méthodologiques qui y sont faits. Aucun spécialiste français ne fait partie du comité d'experts international chargé d'évaluer les réponses à l'appel d'offres de 1997, Paul Dickes, professeur de psychologie différentielle à l'Université Nancy II, n'ayant pas été choisi dans la liste des experts proposés au vote par le Conseil des pays participants (BPC)⁶. La présence de la France dans les instances du programme PISA entre 1995 et 2000 reste incertaine. Si plusieurs membres de la DEP participent régulièrement au Groupe directeur de INES, au Groupe des coordinateurs nationaux de INES et au Réseau A de INES entre 1995 et 1997, si une autre administration centrale du ministère de l'Éducation nationale, la DRIC⁷, représente régulièrement la France au sein du Comité de l'éducation de l'OCDE tout au long de la période, la présence effective de la France reste difficile à établir au sein du réseau INES entre 1998 et 2000 et au sein du BPC et du comité directeur du CERI entre 1995 et 2000 (Bottani & Vrignaud, 2005). La réponse de la France (*PEACE project*, coordonné par l'Iredu de Dijon et mobilisant le réseau de partenaires constitué par la DEP dans les années 1990) à l'appel d'offres de l'OCDE a été classée dernière par le jury international et a obtenu 42 points sur 100 (par rapport à 79 pour le Boston College et 83 pour ACER). À l'exception d'une enseignante intégrée

ponctuellement dans le consortium international concevant PISA 2006, aucun expert français ne participe à cette instance. Par ailleurs, la participation des experts internationaux français n'est pas dictée uniquement par les compétences centrales attendues au sein de chacune d'elles, certains représentants français étant d'abord choisis pour leur maîtrise de l'anglais, leur convergence de point de vue avec les positions de la DEP ou du fait de leur intégration, en tant qu'enseignant, dans un bureau de la DEP chargé de concevoir les évaluations nationales des acquis des élèves en France (Pons, 2009). Cette participation inégale a des implications directes sur le contenu des revendications des représentants et experts français.

C'est la troisième implication politique de la méthode à l'origine de PISA : le rôle de sélection des thèmes et des acteurs légitimes permis notamment par une dichotomie entre les instances à caractère politique (comité directeur, secrétariat général de l'OCDE, groupe des directeurs nationaux de projet) et celles à caractère scientifique (consortium international, groupes d'experts conseils, forum d'experts *ad hoc*). Cette dichotomie concourt à redéfinir des considérations d'ordre méthodologique en considérations politiques, leur prise en compte effective dépendant alors d'autres arbitrages institutionnels. N'ayant pas pu influencer de manière significative l'appel d'offres et la réponse retenue, ne parvenant pas à pénétrer le consortium international, les experts internationaux français portent par exemple la critique des biais culturels au sein du comité directeur de PISA, contribuant ainsi à en faire une question politique et non plus strictement méthodologique⁸. De même, les critiques et propositions méthodologiques récurrentes formulées par les membres français du comité directeur au nom de la DEP (sur le mode de recueil des données relatives à la profession des parents, sur l'hypothèse d'unidimensionnalité ou sur les variables de conditionnement) ont, à leurs dires, fréquemment été ignorées par les membres du consortium ou du secrétariat général de l'OCDE, la réponse d'ordre méthodologique émanant de ces derniers ponctuant la discussion dans une arène dont l'objectif premier n'est pas d'engager des discussions méthodologiques de fond⁹. La capacité des représentants nationaux à infléchir les considérations d'ordre méthodologique dépend donc premièrement de leur capacité à pénétrer les instances où ces considérations vont être perçues comme légitimes. Aux dires de nos interlocuteurs, participer aux groupes d'experts conseils leur a permis de plus faire valoir leur vision des compétences à tester et de la bonne évaluation¹⁰.

La méthode comme outil de contrôle social

Les enjeux politiques suscités par la méthode diffèrent lors de la phase de mise en œuvre dans la mesure où l'activité principale n'est plus d'effectuer des choix méthodologiques mais de faire appliquer ces derniers, même si ces deux aspects se recoupent parfois. La méthode constitue alors un outil de contrôle social, donc l'un des moyens disponibles permettant aux concepteurs et promoteurs de l'enquête de s'assurer de la conformité aux normes des acteurs nationaux chargés d'administrer PISA. La perspective pour ces derniers de ne pas voir figurer les résultats de leur pays dans le rapport final, de les voir figurer avec des amendements méthodologiques stigmatisants, ou de ne pas pouvoir disposer rapidement de la base de données comparatives finale constituée par le consortium pour situer leur pays par rapport à d'autres et produire le document officiel attendu d'eux, sont autant de sanctions possibles. Ces normes ne sont pas uniquement d'ordre méthodologique (mode de collecte des données, modalités de correction des items). Elles se traduisent aussi par une orientation particulière de la production statistique des organismes nationaux, qui se manifeste notamment dans le rythme de production triennal contraignant.

La passation des épreuves PISA en France

Pour s'assurer de la viabilité des tests, l'OCDE prévoit un cahier des charges standardisé. Tout pays ne respectant pas les conditions mentionnées dans ce cahier des charges peut voir ses résultats amendés de plusieurs commentaires, non intégrés à la comparaison, ou même retirés de l'enquête. Ces conditions concernent la traduction des items, les adaptations nationales, l'impression des cahiers d'évaluation, l'échantillonnage, la passation des épreuves, la collecte des carnets de test, les corrections, la saisie des données, leur traitement statistique, etc. Pour la mise en œuvre de ces consignes, des rôles sont définis dans chaque établissement concerné par l'opération : le « coordinateur scolaire » est responsable de la mise en œuvre et l'« administrateur de test » est chargé de la passation des épreuves. En France, ces deux personnes font en général partie de l'équipe de direction de l'établissement (adjoint au chef d'établissement, conseiller principal d'éducation, etc.) mais leur choix est conditionné par certaines recommandations de l'OCDE. Le respect des consignes est vérifié par les maîtres d'ouvrage nationaux (la DEP pour la France) mais aussi par des « moniteurs de contrôle de qualité nationaux » en général proposés par les représentants nationaux, recrutés dans leur

vivier d'experts (principalement des inspecteurs généraux en France) et formés en un temps court par un membre du consortium à la compréhension et la mise en application de la fiche de contrôle qui leur est remise.

La mise en œuvre de PISA en France est de la responsabilité de la DEP, direction d'administration centrale du Ministère, créée en février 1987 sur les bases du service statistique ministériel. Les directeurs nationaux de projet (NPM) sont toujours des membres du bureau de l'évaluation des élèves émanant de cette direction, bureau organisant, depuis le milieu des années 1970, la conception et la passation des évaluations nationales des acquis des élèves. Ces directeurs nationaux sont en général des enseignants de la discipline majeure évaluée dans la session PISA concernée, détachés sur un poste d'administration centrale et maîtrisant l'anglais. L'échantillon des élèves est constitué par l'un des membres de ce bureau qui intègre le groupe de conseillers techniques (TAG) en 2005. Ce bureau est responsable de l'ensemble des aspects logistiques : distribution des cahiers, assistance téléphonique auprès des chefs d'établissement, sélection et organisation des visites des contrôleurs de qualité, organisation de la remontée des données et financement des dépenses de fonctionnement induites par l'enquête et non prises en charge par le consortium. C'est la DEP qui compose également les groupes d'experts nationaux chargés de corriger les items dans chaque discipline et, à une autre étape du programme, de proposer des items que la France pourra soumettre lors des prochaines sessions. C'est elle enfin qui édite la publication officielle du Ministère précisant les résultats de ce pays.

Les formes du contrôle social par la méthode : quelques exemples

En entretien, les représentants français au sein du programme PISA émettent à plusieurs reprises des critiques sur la façon dont l'enquête est supposée être mise en œuvre, critiques qui révèlent en creux les formes du contrôle social exercé à travers la méthode d'enquête, mais aussi les stratégies des acteurs pour résister à ce contrôle, le contourner ou s'en accommoder. Trois exemples en particulier ressortent des entretiens. Le premier renvoie aux critiques méthodologiques adressées à la DEP par le consortium concernant l'échantillonnage des élèves de 15 ans, qui n'intégrerait pas suffisamment les élèves relevant de catégories résiduelles (élèves des DOM-TOM¹¹, élèves scolarisés dans les 3 % d'établissements privés qui n'ont pas contracté avec l'État, etc.) comme en témoigne l'extrait repris¹². Ce dernier illustre bien les formes prises par ce contrôle social (rappel régulier aux obligations méthodologiques définies par le protocole¹³, rappels au professionnalisme attendu

de tout statisticien), ses canaux de diffusion (importance des échanges électroniques, entrée en scène de personnalités du consortium supposées agir comme arguments d'autorité, établissements de conventions sociales suscitant la « confiance » réciproque) mais aussi les modalités de son évitement. La « confiance » peut aussi être mobilisée pour contourner l'impératif méthodologique en toute légitimité. L'affirmation d'une position forte est un moyen de mettre le consortium au défi d'effectivement mettre en œuvre les sanctions à sa disposition, ce qui se fait rarement à coût diplomatique nul.

Le deuxième exemple concerne le refus de la France d'administrer le questionnaire de contexte adressé aux chefs d'établissement, permettant au consortium de lier les résultats obtenus à quelques grandes variables structurelles sur le fonctionnement des établissements scolaires. L'argument technique avancé en entretien est le suivant : l'âge de 15 ans ne correspond pas à la fin de la scolarité obligatoire en France (fixée à 16 ans depuis 1959). À cet âge, les élèves français peuvent être scolarisés dans des établissements très différents selon qu'ils ont redoublé (38% des élèves de 15 ans ont redoublé au moins une fois en France en 2006 comparativement à 13% en moyenne dans les pays de l'OCDE) ou selon qu'ils sont orientés ou non dans une filière générale (en 2008, le taux de réussite au baccalauréat général est de 87,9% par rapport à 77% au baccalauréat professionnel). L'échantillon proposé par la DEP à chaque session de PISA, effectué en deux temps (tirage aléatoire des établissements puis tirage aléatoire des élèves), tient compte de ces particularités (voir le tableau 2). Dès lors, les résultats des différentes catégories d'élèves ne sauraient être, aux yeux de la DEP, représentatifs de l'établissement dans lequel ils sont scolarisés. La très grande majorité des collégiens interrogés dans PISA sont encore collégiens à 15 ans car ils ont redoublé au moins une fois. Leur niveau moyen sera donc significativement inférieur au niveau moyen des collégiens de moins de 15 ans n'ayant pas redoublé. À l'inverse, il est possible que les élèves de seconde interrogés dans PISA aient des résultats globalement supérieurs à la moyenne des élèves de seconde en France dans la mesure où il ne sont composés que d'élèves à l'heure.

Tableau 2
L'échantillon PISA en France : quelques caractéristiques

<i>Session</i>	<i>Structure de l'échantillon</i>	<i>Taux de participation des élèves</i>
2000	4 673 élèves de 15 ans (43 % en seconde générale et technologique, 36 % en troisième, 6,6 % en seconde professionnelle et 10 % en quatrième)	85 %
2003	4 214 élèves de 15 ans (2,2 % en première générale et technologique, 49,6 % en seconde générale et technologique, 34,5 % en troisième, 7,4 % en seconde professionnelle et 5,2 % en quatrième)	50 %
2006	4 700 élèves de 15 ans (2,5 % en première générale et technologique, 49,4 % en seconde générale et technologique, 34,3 % en troisième, 8,4 % en seconde professionnelle, 5,1 % en quatrième)	Non disponible

Source : Pons, 2008, à partir des publications officielles de la DEP.

Ce point pose lui aussi régulièrement problème au consortium aux dires de nos interlocuteurs. Le témoignage en note¹⁴ évoque une forme classique d'incitation à la conformité utilisée par l'OCDE dans ses programmes pour résoudre ce type de particularisme : le recours à la grandeur du pays et au mauvais effet que susciterait sa non-participation ou sa participation non conforme à l'ensemble du programme. Il suggère surtout l'existence de tout un répertoire de stratégies possibles de contournement du contrôle social par la méthode. Le recours au courrier écrit (et non aux courriels réguliers comme dans l'exemple précédent) est un moyen pour la France d'affirmer le caractère solennel de sa contestation ainsi que sa motivation. Le rappel au consortium de ses strictes obligations est un moyen de politiser la querelle méthodologique en réaffirmant l'indépendance et la souveraineté de la France et en renvoyant la résolution éventuelle de la querelle aux arbitrages diplomatiques à effectuer au sein du comité directeur de PISA.

Le troisième exemple concerne le rythme de mise en œuvre de l'enquête PISA. Triennial, il implique un investissement régulier des maîtres d'ouvrage nationaux qui chaque année sont concernés par trois sessions de PISA (expérimentation du prochain PISA, administration du PISA en cours et publication des résultats du PISA précédent). Ce rythme rapide suppose une spécialisation progressive d'une partie du personnel des organismes maîtres d'ouvrage, qui ont le plus souvent bien d'autres missions à honorer. Il contraint les membres du consortium et les maîtres d'ouvrage nationaux à trouver des réponses rapidement à leurs discussions méthodologiques pour ne pas risquer de bloquer toute l'enquête ce qui est plutôt à l'avantage du consortium, légitimé dans ses attentes méthodologiques par la conception même du programme alors que chaque pays prend isolément un risque important en jouant la carte du blocage¹⁵.

La méthode comme élément de structuration du débat public

Comme le notent Lascoumes et Setbon (1996), c'est en grande partie sur la base de la méthode retenue que se fondent la plupart des débats ultérieurs sur une évaluation, que s'apprécie sa pertinence et que se décide son éventuel usage. La méthode permet alors à l'évaluation d'être un moyen puissant de légitimation d'une action politique, celle-ci pouvant se parer de l'objectivité scientifique, de l'efficacité technique ou de la représentativité citoyenne. À l'inverse, elle peut être aussi le prétexte à un déni total des conclusions de l'enquête, parfois sous couvert du bon sens, souvent parce que la dite méthodologie minore ou néglige tel ou tel aspect qui avait en général été mis en avant par les acteurs désireux de participer à la construction du diagnostic, ou parce qu'elle peine à neutraliser tous les effets de contexte porteurs de sens pour les acteurs.

Méthode et argumentaires dans le débat public

L'enquête PISA ne déroge pas à cette règle même si l'effet structurant de la méthode retenue sur le débat public français s'opère à travers des processus de traduction spécifiques (Callon, 1986). Pour comprendre ces derniers, nous avons été conduits dans une recherche effectuée avec Nathalie Mons à proposer la notion d'argumentaire, défini comme un ensemble plus ou moins cohérent et stabilisé d'arguments (scientifiques, techniques ou politiques) qui permettent d'expliquer publiquement les principaux résultats obtenus par les élèves français aux épreuves PISA. Ces arguments peuvent renvoyer aux spécificités de PISA (forme institutionnelle du programme, résultats mis en évidence, méthodologie de l'enquête), aux caractéristiques de l'organisation de l'enseignement en France ou dans d'autres pays (pays modèles ou

repoussoirs) ou faire écho à l'actualité de la politique éducative nationale ou supranationale. Ces argumentaires combinent certains éléments tirés de l'enquête PISA et intègrent ces derniers, avec plus ou moins de justesse et de cohérence, dans un ensemble de considérations antérieures ou parallèles à PISA, relatives à l'histoire, aux identités et aux discours des acteurs qui les énoncent (Mons & Pons, 2010). Si *a priori* tous les acteurs du débat public national sont susceptibles de proposer leur propre argumentaire, l'analyse du débat public suscité en France par PISA entre 1997 et 2007 montre que deux acteurs principaux sont, dès l'origine, en capacité de proposer un argumentaire de portée suffisamment générale pour être potentiellement repris par d'autres et circuler dans le débat public : l'OCDE et le ministère de l'Éducation nationale. Ce dernier propose un argumentaire explicitement fondé sur des considérations méthodologiques, que nous avons appelé « l'argumentaire des biais ».

L'argumentaire des biais : heurs et malheurs (2001-2008)

Cet argumentaire vise principalement à expliquer les résultats de la France à l'enquête PISA par les biais introduits par cette dernière dans l'analyse de ces mêmes résultats. Ces biais renvoient aussi bien à la façon dont a été conçue l'enquête sur le plan international qu'aux caractéristiques de sa mise en œuvre en France. Ils sont de multiples natures : biais culturels, politiques, méthodologiques et même biais de perception (comme lorsqu'on rappelle que PISA n'est en rien une mesure universelle). L'encadré 1 les synthétise, en donnant quelques exemples.

Cet argumentaire se fonde sur les arguments (ou facteurs explicatifs) suivants :

1. *Il existe des biais culturels dans la conception même de l'enquête PISA* (les questionnements sont principalement « anglo-saxons », la France a de meilleurs résultats dans les exercices qu'elle a proposés).
 2. *PISA n'évalue pas des connaissances* (comme le fait la France ou l'IEA) *mais des compétences*.
 3. *PISA n'est pas une mesure universelle mais un outil statistique pratique parmi d'autres* (critique de l'unidimensionnalité et du palmarès final).
 4. *La structure de l'échantillon* des élèves de 15 ans explique la dispersion des résultats et donc leur caractère moyen (voir plus haut dans l'article).
 5. *Les élèves français s'abstiennent de répondre* quand ils ne comprennent pas la question (statut de l'erreur en France, attitude de réserve insuffisamment prise en compte dans l'encodage des items qui valorise de fait ceux qui osent répondre au hasard).
-

Encadré 1. ***L'argumentaire des biais : une version modélisée***

L'argumentaire des biais apparaît, dans sa version complète, dès la première conférence de presse donnée sur PISA en France par la DEP le 4 décembre 2001. Il est présenté principalement par son directeur, en accord avec le cabinet. Il s'appuie également sur les réactions de certains chercheurs qui ont pu collaborer à plusieurs reprises avec la DEP et qui ont plus ou moins directement participé à la conception de PISA sur le plan international. À l'exception de l'argument 3, cet argumentaire est entièrement repris, dans des passages différents, dans la *Note d'information* de la direction publiée à cette occasion. Inégalement repris par les médias, très peu remis en question par les chercheurs et les syndicats d'enseignants, il est diffusé dans sa quasi-totalité par les membres de la DEP à travers plusieurs canaux auprès de spécialistes du champ : séminaire de travail de la FSU¹⁶ (janvier 2002), séminaire de la Direction de l'enseignement scolaire du Ministère sur les comparaisons internationales (mai 2002), dossier spécial de la DEP (novembre 2002), notes internes des inspections générales (novembre 2004), séminaire EHESS¹⁷ d'un inspecteur général (2004-2005), etc.

Le faible développement du débat sur PISA en France entre 2001 et 2004 témoigne en creux du succès que connaît cet argumentaire jusqu'en décembre 2004, date de la deuxième conférence de presse de la DEP relative à PISA 2003, au cours de laquelle les arguments 2 et 4 notamment sont mobilisés. Relativement fermé, cet argumentaire permet en effet d'éviter à tous ceux qui le souhaitent de remettre en question le modèle éducatif français dans un pays qui ne s'est véritablement ouvert aux comparaisons internationales qu'à partir du milieu des années 1990 (Pons, 2008), et de mettre un terme, sur la base d'arguments scientifiques, aux propos de ceux qui au contraire prétendent le faire. L'argumentaire des biais devient progressivement un système d'explication total, se suffisant à lui-même et justifiant qu'on n'approfondisse pas les questionnements soulevés par PISA. Son succès est dès lors paradoxal puisqu'il triomphe dans la mesure où il est faiblement repris.

À partir de 2005 cependant, les mêmes acteurs qui diffusaient l'argumentaire des biais au cours de la période précédente envisagent ceux-ci non plus comme un obstacle rédhibitoire mais comme une invitation à approfondir la réflexion sur la méthode à l'origine de ce type d'enquête. De moins en moins insufflé par la DEP, peu repris dans les médias, l'argumentaire perd progressivement de sa pertinence et de son audience dans le débat public dans un contexte d'accoutumance progressive à l'instrument que constitue PISA, d'un revirement de la part des responsables politiques en matière de

méthodologie à partir de 2006 (à partir de cette date, la position officielle du Ministère n'est plus de critiquer la méthode des comparaisons internationales mais de reconnaître leur validité pour en tirer les enseignements) et d'affirmation progressive d'un autre argumentaire dominant impulsé par les experts de l'OCDE (Mons & Pons, 2010). L'argumentaire des biais réapparaît cependant ponctuellement, dans des conjonctures spécifiques comme en avril 2006 lors de la polémique par déclarations de presse interposées entre Andréas Schleicher, responsable du programme PISA à l'OCDE, et Claude Sauvageot, chef de la mission aux relations internationales et européennes de la DEP, concernant les raisons du refus de la France de publier les résultats de PISA relatifs aux disparités sociales entre établissements. L'argumentaire des biais réapparaît également à partir de décembre 2007 dans les prises de position publiques de différents syndicats, le plus souvent en réaction aux déclarations du Ministre utilisant de plus en plus cet outil. En ayant principalement recours à l'argument 2, il s'agit alors de rappeler que PISA n'évalue qu'une partie de l'enseignement délivré en France et qu'il ne saurait, en aucun cas, justifier les réformes globales du système éducatif annoncées (Mons & Pons, 2010).

Conclusion

L'ambition de cet article était de rappeler le caractère politique des choix méthodologiques effectués dans le cadre des enquêtes internationales des acquis des élèves, au moyen d'une sociologie de l'action publique tentant d'éviter toute perspective normative et dénonciatrice. Il montre que selon les séquences d'action publique retenues, la méthode peut remplir une fonction politique décisive de problématisation de l'action publique, de contrôle social des acteurs concernés par sa mise en œuvre et de structuration du débat public qui se noue autour de ses conclusions. Cette perspective d'analyse constitue à nos yeux une invitation forte à remettre en question les implications politiques des évidences méthodologiques partagées par les acteurs d'un domaine d'action publique donné. Pour cela, elle requiert d'établir des liens qu'on peut supposer aussi délicats que fertiles avec plusieurs pans de la recherche en éducation et en sciences sociales au sens large. Mobiliser la sociologie des examens et des concours permettrait sans doute d'enrichir une sociologie politique des appels d'offres par exemple : ceux-ci ne véhiculent-ils pas une conception élitiste de l'action publique en donnant un avantage aux groupes d'acteurs préconstitués, plus précisément à ceux dont l'activité les a déjà amenés à partager un certain nombre de valeurs, de normes de fonctionnement et de compétences attendues par les commanditaires ?

NOTES

1. Pour une revue de documentation, voir par exemple Mons et Pons (2006). Il conviendrait d'évoquer également les numéros thématiques de revue scientifique consacrés à PISA par le *Scandinavian Journal of Educational Research* (2004), la *Revue française de pédagogie* (2006) ou la *Revista de Educación* (2006), comme de nombreux travaux effectués ou présentés dans le cadre de l'ADMEE et de cette revue.
2. *Programme for International Student Assessment*.
3. Organisation de coopération et de développement économiques.
4. Indicateurs internationaux de l'enseignement. Nous retenons ici comme point de départ l'assemblée générale du projet INES qui s'est déroulée à Lahti en Finlande, en juin 1995, et qui a été l'occasion de proposer un premier projet d'enquête internationale (Bottani & Vrignaud, 2005).
5. Cette période nous permet d'envisager principalement le premier cycle PISA (2000-2003-2006).
6. *International Adult Literacy Survey*.
7. La DEP devient la Direction de l'évaluation, de la prospective et de la performance en 2006. Par commodité, nous n'utiliserons qu'un seul sigle.
8. *Australian Council for Educational Research*. Quatre autres organismes jouent un rôle d'appui auxquels s'ajoutent une quinzaine d'experts européens ou indépendants : CITO, un institut indépendant créé par le gouvernement hollandais en 1968 et spécialisé dans la mesure en éducation au sens large, deux instituts privés nord-américains Westat (organisme composé d'employés actionnaires spécialisé dans l'offre de recherche et d'expertise en statistique et systèmes d'information auprès des agences gouvernementales) et ETS (*Education Testing Service*, organisme ayant une longue expertise dans la constitution de tests d'aptitudes sur tout le territoire américain) et le *National Institute for Education Research* du Japon, spécialisé dans la planification et l'élaboration de politiques éducatives.
9. De manière emblématique, l'appel d'offre international a été adressé à un nombre réduits de souscripteurs éventuels identifiés comme tels par les pays membres de l'OCDE: le Centre international d'étude TIMSS auprès du Boston College à Chestnut Hill (Massachusetts), l'IEA, l'institut hollandais pour le développement des tests CITO, le centre hollandais OCTO, le centre de recherche autrichien pour l'IEA à Salzbourg, ACER et la DEP. Trois réponses ont finalement été adressées au jury international (Bottani & Vrignaud, 2005).
10. *Board of Participating Countries*.
11. Délégation aux relations internationales et à la coopération.
12. «C'est surtout la conception même du protocole d'évaluation, et du coup de la conception de l'éducation [qui posait problème]. [...] Ce qu'on reprochait à PISA, c'était d'évaluer des compétences dans une perspective utilitaire, qui était en gros la perspective du marché de l'emploi.

Par exemple ?

Par exemple... C'est très compliqué d'expliquer ça comme ça. Il faut le faire avec les questionnaires sous la main. [...] Le truc le plus caricatural, c'est de ne pas tester des compétences qui se réfèrent à l'appréciation littéraire d'un texte, mais mettre le paquet sur la lecture d'un horaire de bus ! C'est gros ce que je dis, mais c'est un peu l'idée. Divers types de compétences que nous nous cherchons à enseigner dans notre système éducatif mais qui ne sont pour certains de nos partenaires pas ce qui les intéresse le plus. Eux ils mettent plus l'accent sur l'aspect utilitaire de la lecture, ou pratique, et pas sur ce que l'individu peut en retirer sur l'aspect culturel des choses. Donc les textes ne sont pas ceux sur lesquels on a l'habitude de travailler. À la rigueur pourquoi pas. [...] Mais la malhonnêteté intellectuelle consistait à faire croire que seules ces compétences-là sont utiles, doivent être enseignées, et refuser de tester d'autres compétences qui sont enseignées chez nous, mais pas chez eux ou dans leur vision du monde.

Quelles étaient alors les propositions de la DEP ?

Il y avait des choses très techniques, statistiques comme pour les modèles de réponse à l'item. On a beaucoup travaillé aussi sur les problèmes de traduction. Parce que, imaginez-vous, ils construisent un protocole en anglais qui est ensuite traduit en 45 langues. Qui va aller voir... [...] On a beaucoup insisté sur les biais culturels. Ça c'est notre grand cheval de bataille. » (Entretien avec un membre français du comité directeur de PISA).

13. «Je prends un exemple. Il y a un type d'analyse qu'on monte depuis je ne sais pas combien de temps. Depuis IALS, [...] on fait des classifications de pays selon leur profil de réussite. L'ordre de difficulté est le même pour tous les pays en théorie. Et bien là, on prend les difficultés de chaque pays et on fait une classification des pays selon leur profil de difficultés. Et on s'aperçoit que les pays se regroupent selon des proximités culturelles et linguistiques. [...] Les pays nordiques ont les mêmes points forts et les mêmes points faibles selon les items etc. [...] Donc l'hypothèse d'unidimensionnalité n'est pas vérifiée. Mais ça quand on le présentait au début, ça ne passait pas du tout auprès [du directeur de projet du Consortium PISA] etc. Là je n'étais pas au Tag. C'était la Dep qui avait fait ces enquêtes. [...] Ça, ça a été très mal pris. L'OCDE, [le coordinateur du programme PISA] notamment, a été très agressif là-dessus. En plus, on était étiqueté IALS. [...] Je me souviens lui et [le directeur de projet du Consortium PISA] nous ont écrit par mail en disant que nos analyses n'étaient pas bonnes quoi ! En disant, c'est quoi cette classification. Parce que tout de suite les critiques vont porter sur un point technique. C'est leur truc ça. C'est ce que j'ai compris au Tag. Quand on fait des critiques, c'est toujours sur un point technique que la personne maîtrise. Donc il ne faut pas se laisser bluffer, il ne faut pas lâcher. Donc ils nous critiquaient sur le fait qu'on avait fait la classification sur les taux de réussite aux items et non aux paramètres à la base du modèle selon les items, truc qui finalement aurait rien changé ou très peu aux résultats. Mais pas sur le fond ! Ça a un peu changé, mais PISA 2000 c'était ça ! [...] [Un expert du consortium nommé à titre individuel] a fait cette démonstration sur PISA 2006. C'était super net. Maintenant, ils l'entendent. Ils sont d'accord maintenant pour dire que tout ça est une grosse dimension, un agrégat. Après on peut avoir un débat sur le réalisme des hypothèses de départ [...]. Ce n'est pas simple de toute façon de dire à partir de quand un modèle est unidimensionnel. Mais c'est exactement ce qu'on disait nous en 2000 ! » (Entretien avec un membre français du groupe de conseillers techniques).

«Un autre point d'achoppement, ça a été le recueil de la profession des parents. Après l'expérimentation de 1999, avec l'équipe méthodologique de l'INSEE, on a appelé [au téléphone] un sous-échantillon des familles dont les enfants avaient participé à l'expérimentation pour encoder directement la PCS sur la base des questionnaires de l'INSEE. C'était une prise directe d'information. Et bien on a observé des divergences fortes entre ce que déclarent les élèves et les professions telles que nous les avons directement encodées. Donc la DEP n'a pas refusé cette variable, mais on voulait que notre enquête soit prise en compte pour améliorer le processus de recueil de l'information sur cette variable. Au début [...] ils étaient super rigides. Ils ne voulaient pas entendre parler de cette enquête. Ils étaient assez agressifs là-dessus. En gros, on avait demandé aux parents directement mais les parents ne savent pas eux-mêmes ! Ils avaient des arguments comme ça. Ils avaient recalculé ça en indice en disant que globalement, il y avait peu d'impacts sur l'indice. Ils étaient complètement dans le déni et le refus. En plus, on n'était pas les seuls. Il y avait les Britanniques qui avaient produit une étude et la République Tchèque je crois. C'était en 1999-2000. Et ils montraient la même chose que nous, avec des taux de convergence entre la réponse des parents et celle des enfants assez proches des nôtres.» (Entretien avec un membre français du groupe de conseillers techniques).

«Dans leur modèle, ils ont des variables de conditionnement. Ils en ont besoin pour estimer les scores de tout un tas de variables pour – je te la fais rapide – pour avoir une idée des distributions dans chaque pays, pour avoir une meilleure estimation. [...] Elles sont utilisées après, pour réfléchir sur l'équité, mais elles sont utilisées aussi pour calculer les scores, ce qui peut paraître étrange. [...] Toutes ces variables-là, ACER est rigide là-dessus, parce qu'il ne veut pas les changer. Il dit que si on les change, ça va changer complètement les résultats du modèle. Je ne suis même pas sûr que ça ait été vérifié un jour. Donc on a tout un corpus de variables comme ça qu'on n'a pas le droit de modifier. Là on atteint une limite très claire d'où commence l'intérêt d'Acer et s'arrête celui de l'investigation scientifique.» (Entretien avec un membre français du groupe de conseillers techniques).

14. **«Quel a été votre message principal dans les groupes PISA ?**

Par exemple en science. La manière dont le premier cycle PISA abordait la culture scientifique, c'était uniquement par la lecture d'articles de journaux. Pour nous, c'était plus une analyse des compétences en lecture que l'évaluation de compétences scientifiques. On a réussi par le biais d'un expert dans le groupe à ce que le choix se porte sur des articles plus scientifiques et qu'on évalue le réinvestissement de compétences scientifiques dans un cas pratique.» (Entretien avec un chef du bureau des évaluations des élèves de la DEP, membre du réseau A INES).

«On a trouvé un certain nombre de biais [dans l'enquête IALS], ce qui nous a permis après, moi en tête, d'être très vigilants sur les traductions de PISA et de demander des révisions de traduction de PISA. Et j'ai été le principal déclencheur. J'ai été suivi mais j'étais le déclencheur de la révision des traductions. Et on m'en a... C'est vrai que le consortium n'était pas très satisfait de la chose.

Que vous vous manifestiez ou de la traduction ?

Non qu'on se manifeste. [...] Les Australiens qui ont peu de connaissances des langues étrangères et qui sont dans une langue dominante, pour eux la question de la traduction n'est pas une question pertinente. Même s'ils doivent répondre aux pays en disant qu'ils répondent au mieux, la finesse qu'il peut y avoir dans la mise en mots... Et [un expert du consortium nommé à titre individuel], qui s'est occupée de l'agence de traduction (ça ne s'appelait pas comme ça), en tout cas n'était pas du tout sensible aux arguments. [...] Quand je suis allé à Melbourne pour la première fois, j'ai découvert vraiment ce monde de PISA, ce monde des Australiens alliés avec Cito qui n'avaient pas de document de base sur la lecture, si ce n'est, si ce n'est – et c'est là que c'est intéressant – l'enquête IALS et toutes les enquêtes qui ont pu être conduites du côté américain par ETS, que ça soit NAEP ou autres. Et les enquêtes conduites par StatCan. C'était leur vivier. Ils ne se posaient pas de question, ils allaient adapter ces outils même si IALS est une enquête pour les adultes.» (Entretien avec un membre français du groupe d'expert conseil en compréhension de l'écrit).

15. Départements et Territoires d'Outre-Mer.

16. «Par exemple les TOM sont exclus parce que bon souvent les programmes sont décalés, le temps scolaire est décalé aussi avec les vacances. Et puis bon ça coûte cher. Aller à Saint-Pierre et Miquelon pour évaluer trente gamins, ça coûte cher. Voilà il y a toute une partie comme ça de la population qu'on élimine. [...] Ça, rien que pour exclure ces catégories, il y a toute une discussion avec le consortium. [...] Eux naturellement, ils cherchent à en inclure un maximum. Nous, pour des raisons pratiques, on espère éliminer certaines populations. [...] Et là pour 2009, ben voilà, ils reviennent à la charge tous les trois ans sur comment on définit l'échantillon. Là ils nous ont titillés sur les privés hors contrat qu'on exclut, alors que c'est particulier. La plupart du temps, ils ne souhaitent pas faire les évaluations du Ministère, c'est tellement peu d'élèves. Dans quelle strate tu mets ces privés hors contrat? Faut-il créer une strate pour le privé? C'est compliqué. On a eu tout un débat par mails. [Le Directeur du Consortium PISA pour le traitement et l'analyse des données et pour le contrôle de la qualité] s'en est mêlé. Moi je considère que c'est un détail. Ils pèseront en rien dans l'enquête. Et en plus c'est une question pratique. Mais bon on passe beaucoup de temps comme ça à remettre sur le tapis des questions qui avaient été actées trois ans auparavant. [...] Ils remettent à plat à chaque fois.

Comment faire pour les convaincre du retrait de certaines catégories? Parce que je me mets à leur place, si on retire toujours certaines catégories, comment savoir qu'elles n'ont pas d'effet sur l'échantillon...

Oui tout à fait. Après c'est une question de confiance aussi. On se connaît. Il y a une confiance ou pas. Après ça se joue... ça s'impose aussi. Il y a un moment où il faut savoir dire: "non on fera comme ça. Stop avec ces débats qui vont nous prendre la semaine en échanges de mails!". Ça va rarement jusqu'à une réaction de leur part... Après qu'est-ce qu'ils peuvent faire? Oui ils mettent des avertissements dans les rapports, mais bon nous on n'a jamais eu ce problème.» (Entretien avec un membre français du groupe de conseillers techniques).

17. D'après ce dernier, il n'est pas possible par exemple d'exclure des établissements de l'échantillon pour cause de mauvais fonctionnement au-delà d'un certain pourcentage (par exemple pas plus de 0,5% de l'échantillon pour PISA 2006), ni d'exclure dans ces établissements certains élèves du fait de leurs particularités au-delà d'un certain pourcentage (par exemple PISA 2006 prévoit que le pourcentage d'élèves exclus pour cause de handicap fonctionnel ou intellectuel ou de maîtrise insuffisante de la langue du test ne peut dépasser 2%).

18. « Nous refusons de faire passer le questionnaire sur les établissements. Là on a fait un courrier. Point. On essaie aussi de remettre les choses en place. Le consortium, il fait l'enquête mais ce n'est pas lui qui gère le projet. Ça, on le décide avec l'OCDE.

Parce que cette pression, tu la ressens ?

Sur le questionnaire établissement ? Oui ils reviennent à chaque fois : " comment, un pays comme la France qui ne passe pas le questionnaire..." » (Entretien avec un membre français du groupe de conseillers techniques).

19. « Est-ce qu'on a besoin d'un PISA tous les trois ans ? Je trouverais une période de cinq ans plus significative. Les résultats seraient digérés. Ça correspondrait plus au rythme des systèmes éducatifs. Mais trois ans, c'est un rythme politique, à la fois du point de vue de l'OCDE et des pays. Les pays peuvent communiquer toujours du PISA. Et pour l'OCDE, les pays font tout le temps PISA. [...] Nous on a tout le temps un PISA en cours. Le dossier de PISA 2006 n'est pas sorti, il a fini d'être rédigé il n'y a pas longtemps. Bon on est dans PISA 2009 depuis deux ans. Et PISA 2012 commence, avec l'appel d'offre. Donc on est tout le temps dans PISA. D'un point de vue de gestion, c'est malin aussi. On a toujours la même charge de travail sur PISA [...]. Donc pour les pays, il faut dédier toujours des moyens à ça. » (Entretien avec un membre français du groupe de conseillers techniques).
20. Fédération syndicale unitaire.
21. École des Hautes Études en Sciences Sociales.

RÉFÉRENCES

- Bézès, P. (2009). *Réinventer l'État*. Paris : PUF.
- Bottani, N., & Vrignaud, P. (2005), *La France et les évaluations internationales des acquis des élèves*. Rapport établi à la demande du HCEE, mai 2005.
- Callon, M. (1986). Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques et des marins pêcheurs dans la baie de Saint-Brieuc. *L'année sociologique*, 36, 169-208.
- Callon M. (éd.) (1989). *La science et ses réseaux. Genèse et circulation des faits scientifiques*. Paris : La Découverte.
- Callon, M., & Latour, B. (éds) (1985). *Les scientifiques et leurs alliés*. Paris : Pandore.
- Carvalho, L. M., & Costa, E. (2009). *Production of OECD's "Programme for International Student Assessment" (PISA) Rapport pour le projet Know&Pol*, Université de Lisbonne, Faculté de psychologie et des sciences de l'éducation.
- Crossley, M. (1999). Reconceptualising comparative and international education. *Compare*, 29(3), 249- 267.
- CURAPP (1993). *L'évaluation dans l'administration*. Paris : PUF.
- Duran, P. (2002). L'évaluation au risque de son institutionnalisation. *Politiques et management public*, 20(4).
- Duran, P., & Monnier, E. (1992). Le développement de l'évaluation en France : nécessités techniques et exigences politiques. *Revue française de science politique*, 42(2), 235-262.
- Goldstein, H. (1995). *Interpreting international comparisons of student achievement*. Paris : UNESCO Publishing.
- Grek, S. (2009). Governing by Numbers: the PISA "effect" in Europe. *Journal of Education Policy*, 24(1), 23-37.
- Harris, D. N., & Herrington, C.D. (2006). Accountability, standards, and the growing achievement gap: lessons from the past half-century. *American Journal of Education*, 112, 209-238.
- Henry, M., Lingard, B., Rizvi, F. & Taylor, S. (2000). *The OECD, Globalisation and Education Policy*. Oxford: Pergamon Press.
- Jacob, S., & Varone F. (2004). Institutionnalisation de l'évaluation et Nouvelle Gestion Publique : un état des lieux comparatif. *Revue internationale de politique comparée*, 11(12), 271-292.
- Lascoumes, P., & Le Galès, P. (2004), *Gouverner par les instruments*. Paris : Les Presses de Sciences-Po.
- Lascoumes, P., & Setbon, M. (1996). *L'évaluation pluraliste des politiques publiques. Enjeux, pratiques, produits*. Cachan : GAPP-CNRS.
- Lawn, M. (2006). Soft governance and the learning spaces of Europe. *Comparative European Politics*, 4(2), 272-288.
- Maroy, C. (2006). *École, régulation et marché. Une comparaison de six espaces scolaires en Europe*. Paris : PUF.
- Mollis, M., & Marginson, S. (2002). The assessment of universities in Argentina and Australia: Between autonomy and heteronomy. *Higher Education*, 43(3), 311-330.

- 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), 425-435.
- Mons, N. (2007b). *Les nouvelles politiques éducatives*. Paris : PUF.
- Mons, N. (2008). Évaluation des politiques éducatives et comparaisons internationales. *Revue française de pédagogie*, 164(3), 5-13.
- Mons, N., & Pons, X. (2006). *Les standards en éducation dans le monde francophone : une analyse comparative*. Neuchâtel : IRDP.
- Mons, N., & Pons, X. (2010). The reception of PISA in France: a cognitive approach of institutional debate (2001-2008). *Sisifo Educational Sciences Journal*, janvier, disponible sur [http://sisifo.fpce.ul.pt].
- Morgan, C. (2007, septembre). *The OECD Programme for International Student Assessment: Unraveling a Knowledge Network*. Thèse doctorale, School of Public Policy and Administration, Université Carleton, Ottawa, Ontario.
- Normand R. (2010). Expertise, networks and indicators: the construction of the European strategy in education. *European Educational Research Journal*, 9(3), 407-421.
- Novoá, A. (2001). État des lieux de l'éducation comparée : paradigmes, avancées et impasses, In R. Sirota (éd.), *Autour du comparatisme en éducation* (p. 41-68). Paris : PUF.
- OCDE, (2002). *La lecture, moteur de changement. Performances et engagement d'un pays à l'autre. Résultats du cycle d'enquêtes de PISA 2000*. Paris : OCDE.
- Papadopoulos, G. S. (1994). *L'OCDE face à l'éducation : 1960-1990*. Paris : OCDE.
- Pons, X. (2008). *L'évaluation des politiques éducatives et ses professionnels. Discours et méthodes (1958-2008)*. Thèse de doctorat, Sciences-Po, Paris.
- Pons, X. (2009). *The French contribution to the international design of PISA*. Document de travail pour le projet Know&Pol. Paris: OSC, février 2009.
- Pons, X. (2010). *Évaluer l'action éducative. Des professionnels en concurrence*. Paris : PUF.
- Schwartz, O. (1993). Postface. L'empirisme irréductible. In N. Anderson, *Le hobo : sociologie du sans-abri* (p. 265-308). Paris : Nathan.
- Spentheuer, V. (1998). *L'évaluation des politiques publiques, avatar de la planification*. Thèse de doctorat, Université Pierre-Mendès-France, Institut d'études politiques de Grenoble.
- Vrignaud P. (2005). L'INETOP, 75 années dans l'histoire de l'évaluation psychologique et pédagogique en France. *L'Orientation scolaire et professionnelle*, 34, hors-série, 171-197.
- Vrignaud P. (2006). La mesure de la littéracie dans PISA : la méthodologie est la réponse, mais quelle était la question? *Revue française de pédagogie*, 157, 27-43.
- Weber, M. (1965). *Essais sur la théorie de la science*. Paris : Plon.

Date de réception : 11 juin 2011

Date de réception de la version finale : 5 septembre 2011

Date d'acceptation : 8 septembre 2011

