Édito - Les enjeux des évaluations internationales

Faouzia Kalali et Edgar W. Jenkins

Édition électronique

URL : <https://journals.openedition.org/ree/5330>

DOI : 10.4000/ree.5330

ISSN : 1954-3077

Éditeur

Université de Nantes

Référence électronique

Faouzia Kalali et Edgar W. Jenkins, « Édito - Les enjeux des évaluations internationales », *Recherches en éducation* [En ligne], 14 | 2012, mis en ligne le 01 septembre 2012, consulté le 09 juillet 2021. URL : <http://journals.openedition.org/ree/5330> ; DOI : <https://doi.org/10.4000/ree.5330>

Recherches en éducation est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Édito

Les enjeux des évaluations internationales

Les évaluations internationales influencent fortement depuis plusieurs décennies le paysage des politiques de l'enseignement dans la plupart des pays du monde¹. La première en date est celle de l'IEA (International Association for the Evaluation of Educational Achievement) qui, depuis 1958 jusqu'à ce jour, organise régulièrement des enquêtes en littéracie (PIRLS), en sciences et en mathématiques (TIMSS). Depuis 2000, l'OCDE (Organisation de Coopération et de Développement Economiques) a lancé sa propre enquête PISA, renouvelée tous les trois ans, pour évaluer la culture mathématique, scientifique et la compréhension de l'écrit. A ce titre, les résultats de ces enquêtes telles que PISA (Programme international pour le suivi des acquis des élèves), TIMSS (Trends in International Mathematics and Science Study) ont influencé les réformes curriculaires en Europe, au Canada, en Australie et partout dans le monde. Le Centre pour la recherche et l'innovation dans l'enseignement, le CERI de l'OCDE, ou encore le Bureau international d'éducation, le BIE de l'UNESCO, contribuent au suivi de ces réformes curriculaires dans différents pays du monde. Des enjeux économiques et sociaux façonnent actuellement l'éducation, en France et à des échelles européenne et internationale : la rhétorique porte sur les connaissances et les compétences nécessaires à tous les jeunes abordant leur vie d'adulte, quelle que soit la place qu'ils occuperont dans la société et quel que soit leur désir d'avenir (Kalali, 2010).

Les enquêtes internationales ont produit un ensemble sans précédent de données quantitatives complexes sur les performances des élèves de différents âges dans les différents systèmes éducatifs, et sur les facteurs qui influencent ou déterminent ce rendement. Il existe également un ensemble de recherches ou d'études associées qui accompagnent cette masse de données. Certaines d'entre elles se sont attachées à identifier à partir des données disponibles toute politique d'éducation ou autres retombées ; d'autres cherchent à questionner différents aspects, y compris leur méthodologie et les techniques statistiques qu'elles emploient dans leur analyse. Il est admis que les résultats de ces enquêtes ont réussi à mettre au premier plan la performance relative des systèmes éducatifs, des écoles et des différents groupes d'élèves, et à attirer l'attention sur les facteurs jugés significatifs en expliquant les différences ou les écarts qui peuvent exister (Baudelot & Establet, 2009). Ces résultats sont, par conséquent, principalement d'intérêt pour les décideurs et les chercheurs, bien qu'ils présentent également des questions importantes pour les enseignants et les autres professionnels.

Comparaisons internationales : comment utiliser les résultats ?

Les comparaisons internationales et les recherches qui les sous-tendent inaugurent une nouvelle organisation de la prise de décision politique et de la production scientifique en matière d'éducation. Les analyses dominent le paysage actuel à divers niveaux (politiques, enseignants, parents, élèves, institutions scolaires, programmes et curriculums, sphère publique...). Cette masse de données bute sur des obstacles comme la diversité des acteurs, l'absence de

¹ L'étude pilote (Foshay, 1962) à laquelle ont participé douze pays émane d'un groupe d'éducateurs de haut niveau qui a pris l'habitude de se réunir à l'Institut de l'UNESCO pour l'éducation, à Hambourg, au sein duquel on comptait G. Mialaret.

consensus et la dominance de modèles anciens (Jenkins, 2006). Et les recherches n'ont pas conduit à des analyses pertinentes à tous les niveaux.

On peut dire que la plupart de ces enquêtes reflètent une position inductive : collecte de données et analyse des résultats, principalement en termes de classement, de percentile et de corrélation. Elles génèrent quelques hypothèses sur la façon dont les données internationales peuvent contribuer à expliquer le rendement des systèmes éducatifs et des groupes d'élèves. Sont beaucoup moins fréquentes les études internationales qui adoptent une approche hypothético-déductive, dans laquelle une hypothèse spécifique ou différentes hypothèses sont testées et qui, dès le départ, déterminent la conception globale, l'analyse et les essais. L'accent mis dans les publications sur les échelles de scores et les classements des pays est un corollaire de la première approche.

Un autre problème concerne le manque de prudence des professionnels de l'éducation et des politiques lorsqu'ils ne parviennent pas à comprendre que les différents systèmes éducatifs doivent être considérés à un niveau macro au sein de modèles sociaux, économiques et politiques complexes. Ceux-ci doivent tenir compte également des valeurs culturelles incarnées. Compte tenu de cette complexité, il semble peu probable que les politiques et les pratiques qui caractérisent un système d'éducation pourraient jamais être transférées de façon simple dans un autre système éducatif. Néanmoins, cette limitation reçoit rarement l'attention qu'elle mérite. De même, le renouvellement de telles enquêtes à des périodes différentes offre la possibilité d'identifier des tendances dont l'intérêt est certes primordial pour les décideurs et tous ceux qui s'intéressent aux systèmes éducatifs, aux écoles et aux groupes d'élèves. Les données recueillies ne peuvent nullement constituer une estimation dans le long terme de l'action des facteurs qui influencent l'apprentissage. Les apports des enquêtes internationales comme PISA, TIMSS ou encore PIRLS peuvent être évalués plus rigoureusement en tenant compte de ces limites.

Dans ce dossier, des contributions détaillent les préconisations qui sont faites dans le cadre des études comme TIMSS, PIRLS et PISA. Ecrites par ceux qui sont directement impliqués dans ces enquêtes, l'approche adoptée est en contraste avec la position indépendante et critique de ceux qui sont impliqués dans la recherche. Néanmoins on peut relever quelques convergences comme l'utilisation sélective des résultats qui renvoie au problème de l'interprétation des données. Enfin une réflexion philosophique plus générale sur les finalités de l'école mérite l'attention car elle remet au centre la question des enjeux.

Ce que préconisent les enquêteurs à l'attention des politiques

Pour mener des analyses en termes de corrélations, on a besoin de la mise en place d'hypothèses fortes, testées à l'aide d'un dispositif expérimental rigoureusement conçu. Dans leur article Maciej Jakubowski et Francesca Borgonovi, analystes de l'OCDE, mettent en lumière diverses données de PISA². Leur article illustre les limites de la posture hypothético-déductive, car il souligne les difficultés de mener des analyses en termes de corrélation à partir des données croisées des divers pays. Se basant sur une combinaison d'hypothèses, à partir de résultats inédits de l'enquête de 2009, Maciej Jakubowski et Francesca Borgonovi mettent en garde les décideurs qui cherchent à tout prix à établir l'impact des politiques d'éducation sur les performances des élèves. Sur ce point, Mullis et Martin voient en ce qui concerne TIMSS et PIRLS, la nécessité d'évaluations ultérieures pour apprécier l'impact des mesures prises à la lumière des problèmes révélés par les enquêtes. Toutefois, la difficulté est de s'assurer que ce sont les impacts des réformes – et aucun autre facteur – qui sont à l'origine des modifications mesurées. On pense que les analyses de corrélations des données peuvent suggérer des relations entre différentes variables, mais leur pouvoir explicatif réel peut venir uniquement des

² L'OCDE lançait son deuxième cycle d'enquêtes PISA en 2009 avec de nouveau la compréhension de l'écrit en tant que domaine majeur d'évaluation comme en 2000, l'étape suivante en 2012 donnera le rôle majeur aux mathématiques comme en 2003, et enfin en 2015 les sciences seront le domaine majeur comme en 2006.

enquêtes longitudinales. Ces dernières sont particulièrement difficiles à concevoir et à mettre en œuvre, particulièrement dans un contexte international (taux d'érosion souvent élevé et problème de conservation d'échantillons adéquats particulièrement grands dans certains pays les plus pauvres du monde). Tirer des conclusions à partir des résultats issus de différentes périodes exige l'hypothèse de la comparabilité des échantillons dans le temps, et de la représentativité de la population sous-jacente. Les analystes de l'OCDE en font l'illustration à travers leur article qui détaille un exemple fondé sur des différences entre les sexes. Mais là aussi, on pense qu'il est nécessaire de supposer que tous les facteurs envisagés affectent garçons et filles dans la même mesure et de la même façon. Ces hypothèses sont trop souvent négligées ou difficilement validées. L'article donne, au prix d'analyses statistiques très poussées, inspirées de l'économétrie moderne, une explication causale des écarts de performance entre les sexes en compréhension de l'écrit. L'exemple choisi est opportun, car il montre que la dimension du genre se révèle être un levier à géométrie variable. La plupart des filles rattrapent leur retard sur les garçons en science et en mathématiques. Les garçons n'y arrivent pas en lecture. Quelles retombées ? L'article précise que les choix politiques devront viser à encourager les filles les plus performantes à étudier les mathématiques et les sciences aux niveaux les plus avancés. Pour les garçons, les politiques visant à aider à l'acquisition de compétences en compréhension de l'écrit sont tout particulièrement nécessaires pour ceux qui présentent des difficultés de lecture. Car la faiblesse du niveau de compétences en compréhension de l'écrit parmi les garçons les moins performants est encore plus préoccupante.

Ina Mullis et Michael Martin, directeurs exécutifs de TIMSS, adoptent dans leur contribution une approche quelque peu différente pour les résultats de TIMSS et PIRLS. On note que ces programmes sont l'occasion pour certains pays en développement de recueillir des données sur leurs systèmes éducatifs. Néanmoins on peut se demander si les impacts sont uniformément bénéfiques et questionner ainsi la légitimité des bonnes pratiques. Les auteurs reconnaissent que ces études servent d'impulsion aux politiques en matière d'éducation. TIMSS et PIRLS sont assez souvent l'occasion pour les pays de lancer des réformes des programmes scolaires et de pédagogie. Mais on peut penser que l'intention est plus de gagner en place dans les classements des pays que d'engager des transformations profondes. Ces enquêtes montrent également la façon dont les résultats éclairent les différences régionales dans le rendement des élèves dans un système d'éducation donné, et attirent l'attention sur ces différences qui sont parfois plus grandes que celles qui existent entre les différents systèmes éducatifs. Cet aspect est souvent masqué dans les comparaisons internationales au profit des classements globaux entre pays. Il y a ici un agenda pour les chercheurs qui souhaitent autant expliquer les écarts à l'intérieur des pays qu'entre eux et, ce faisant, identifier les divers liens de causalité entre les facteurs impliqués. La mise en garde contre le risque de l'utilisation sélective de données, qui servent à légitimer des décisions politiques, est ainsi affirmée dans les deux textes. Ce risque interpelle les chercheurs qui souhaitent s'assurer que les décideurs sont bien conscients de la nécessité de faire preuve de prudence dans l'élaboration de politiques fondées sur les conclusions des études TIMSS, PISA et PIRLS.

Une lecture critique émanant des chercheurs

Les apports des enquêtes internationales posent donc le problème de l'articulation de la recherche et de l'action politique (Jenkins, 2000 ; Olsen & Lie, 2006) : chercheurs et décideurs travaillent à des échelles de temps différentes. Les législateurs peuvent faire peu de cas des conseils qui ne leur donnent pas une ligne de conduite sans ambiguïté, fondée sur des preuves ou qui leur propose des alternatives. En outre, les chercheurs ont besoin d'être sûrs, en ce qui concerne les demandes des politiques, que ceux-ci ne prennent pas trop de paramètres comme « donnés » et qu'ils n'ignorent pas les questions plus fondamentales qui *peuvent* être d'importance dans le long terme.

Nicolas Coppens propose dans son article d'analyser les résultats des élèves français au sujet de la culture scientifique. Basé sur les données du PISA 2006 et 2009, l'article se situe dans le contexte scolaire français (changements de programmes, réforme du socle commun...). Cette

évaluation paraît spécifique en ce sens que les écarts entre les sexes sont très faibles par rapport aux évaluations qui portent sur la culture mathématique ou la lecture (souligné également par l'OCDE ci-dessus). On peut penser aux enjeux de compétitivité qu'induit l'évaluation différenciée des domaines comme la lecture, les mathématiques et les sciences. Ces dernières continuent à alimenter les propos alarmistes depuis les rapports d'expertise sur la chute des effectifs et la situation défavorable des filles (UE, 2004). Si l'on examine les résultats fournis dans l'article, on voit que 8 % des élèves français atteignent en science les niveaux 5 et 6. La proportion des élèves à faible rendement est élevée, 20 % des élèves n'atteignent pas le niveau 2 qui correspond au seuil à partir duquel on peut être considéré comme capable de participer de manière efficace et productive aux débats qui agitent la société. Ces données nous interpellent quand on pense que les nouveaux programmes scientifiques au collège insistent davantage sur une culture scientifique commune. Et on ne peut que constater le désarroi qui peut s'emparer des enseignants. Dans le même temps, on a assisté assez récemment à une prise en compte de la disparité entre les groupes d'élèves au sein des pays. L'article de l'OCDE, dans ce dossier, signale d'emblée que PISA identifie les pays qui réussissent à obtenir à la fois un niveau de performance élevé pour leurs élèves et une répartition équitable des possibilités d'apprentissage. Pour notre part, nous avons relevé un certain nombre de publications qui vont dans ce sens. En 2009, l'OCDE publie *Top of the class, high performers in science in PISA 2006*. Les données sont celles de PISA 2006. Ce nouveau volume exploite principalement les données en relation avec quatre groupes d'élèves de différentes performances³. Ce volume est à examiner en parallèle avec une autre publication de l'OCDE. Il s'agit de *Regards sur l'éducation* de 2009 qui détaille les indicateurs de l'OCDE : indicateur A4 (le profil des élèves les plus performants en science à 15 ans), indicateur A5 (les attitudes et les motivations à l'égard des sciences des élèves les plus performants en science dans PISA 2006). En plus de ces méta-analyses, nous attirons l'attention sur l'analyse faite par Baudelot et Establet (2009) qui cible le problème des élèves en difficulté. Notre hypothèse pour expliquer cette prise de conscience est qu'il fallait attendre la fin du premier cycle de PISA pour disposer d'un panorama complet des trois domaines, sur le plan des compétences. Il faudra sans doute attendre la fin du deuxième cycle (2015) pour évaluer les effets des réformes en science entreprises à la lumière du PISA. Réalisant qu'il n'y a pas de différence entre les résultats de 2006 et ceux de 2009, l'auteur souligne à juste titre qu'il ne peut pas tirer de conclusions. Si les programmes actuels font la promotion de certaines compétences en adéquation avec la définition de la culture scientifique telle qu'elle ressort du PISA, il est encore trop tôt pour évaluer les retombées de la réforme du socle commun des connaissances et des compétences. En France, les nouveaux programmes de collège ont connu plusieurs remaniements. Dans d'autres pays comme la Norvège (texte de Sjøberg), ce sont les ministères et les politiques d'éducation qui ont subi des remaniements.

Svein Sjøberg présente une critique du PISA appuyée sur de multiples rapports de l'OCDE et se référant au cas de la Norvège. Déclarant que PISA est un projet politique et non une démarche critique, il démonte, étape par étape, ce qu'il appelle l'*entreprise* PISA et cible des aspects problématiques jusqu'ici négligés par la recherche. Face à une forte demande sociale, à des sociétés en constante mutation, nous constatons que les évaluations supranationales montrent une vision prospective à l'œuvre dans les réformes de socles communs de connaissances et de compétences qu'elles influencent. On a donc un contexte peu ordinaire où les missions de l'éducation scientifique s'élaborent à une échelle internationale. Cette internationalisation des enjeux et le traitement global des questions éducatives induisent des logiques d'inclusion, et permettent la promotion des bonnes pratiques qui se font au détriment des contextes locaux, et nationaux (Kalali, 2010). Justement l'auteur attire notre attention, à l'appui d'arguments économique et politique souvent ignorés des chercheurs, sur l'écart qu'il y a entre ce que l'on estime évaluer et le statut que l'on accorde aux résultats obtenus. Manifestement, il s'insurge contre la vision globalisante que véhicule PISA et débusque ainsi les intentions cachées qu'est par exemple l'uniformisation des systèmes éducatifs des différents pays.

³ Par exemple, les élèves les plus performants en sciences sont ceux qui se situent aux niveaux de compétence 5 et 6. Les élèves performants sont définis comme étant ceux qui sont au niveau 4, les élèves moyennement performants sont aux niveaux 2 et 3, et les élèves moins performants se situent au niveau 1 ou en dessous.

Si l'auteur souligne la nécessité et l'urgence d'une recherche indépendante, et met en garde contre l'utilisation sélective des données du PISA pour des recommandations politiques, il ne montre pas les voies par lesquelles on peut réengager le débat démocratique pour la réflexion collective sur le projet de l'école. Normand (2003), qui voit ce dernier en tant que projet politique, souligne la nécessité de création de lieux de concertation publique à l'échelon européen ou international pour que s'exprime la diversité des intérêts dans le respect de principes universels concernant l'éducation. Sjøberg souligne à plusieurs reprises que PISA, en allant au-delà des programmes nationaux, ignore et fait passer à la trappe la question des finalités dans les débats publics et dans les discours des politiques. Serait-il d'accord pour une discussion des finalités (actuellement "missions" sur le plan politique) qui doit être aussi le fait des citoyens (Martinand, 1989) ? Dans tous les cas, nous pensons que de telles concertations publiques ne peuvent se faire à un échelon supranational sans gommer les intérêts locaux. La mesure de la qualité du système éducatif des pays à l'aide d'indicateurs universels est une mission impossible selon l'auteur.

Quelles réflexions sur les finalités de l'école ?

Nous avons vu que les préconisations peuvent revêtir un caractère formel, la position indépendante et critique qu'adopte la recherche permet d'envisager un certain nombre de possibles. Reste la question épineuse des finalités de l'école. Tout le monde convient que les enquêtes internationales ont impulsé des réformes dont les enjeux rejoignent les citoyens dans leur vie quotidienne. Une discussion des finalités est souvent escamotée au profit d'une doxa administrative que sous-tend une rhétorique managériale (Kalali, 2010). L'intérêt du texte d'Alain Trouvé est qu'il cible, dans son examen de la crise de l'école, l'idéologie néolibérale qui s'impose à l'éducation via les directives européennes et l'emprise marchande exercée par les psychotechnologies sur la jeunesse. Plus qu'un glissement sémantique, l'auteur nous montre que ce sont en fait des intentions, des actes et des pratiques qui participent à la « délégitimation » de l'école dans ses valeurs et ses orientations classiques. Son examen de la situation en France montre une rhétorique managériale qui rejoint la *nouvelle gestion publique* que dénonce Sjøberg dans son analyse de l'école en Norvège. Dès lors, on peut se demander si la question des finalités repose essentiellement sur la mise en avant des valeurs comme le pense l'auteur. Ces dernières sont étrangères aux logiques technocratiques des divers experts. Ainsi est posé le problème d'une prise en charge par le curriculum (Fensham, 2007). La sortie de crise qu'esquisse l'auteur se centre sur des enjeux de réappropriation culturelle des savoirs par l'école, une réappropriation qui passe par les contenus, les attitudes et les méthodes d'enseignement. Elle est directement politique et pas seulement pédagogique, affirme l'auteur.

Un double agenda ?

L'importance accordée aux enquêtes internationales et l'intérêt que leur portent les décideurs ont transformé les politiques éducatives. Tous les articles de ce dossier montrent que l'école est une affaire de politique et pas seulement de pédagogie. Nous avons rappelé que dans de nombreux cas, il est peu probable que les préconisations, qui émanent des organismes intergouvernementaux comme l'OCDE ou des institutions de recherche comme l'IEA, constituent une réponse simple susceptible d'éclairer les législateurs et les décideurs. Cependant, les responsables politiques voudront toujours expliquer les questions de la réussite et de l'apprentissage sans tenir compte du facteur temps. Dans quelle mesure les évolutions sont à attribuer à des facteurs sociaux plus larges, pour nous permettre de délibérer au sujet des interventions politiques ayant une incidence sur les programmes scolaires et la pédagogie ? L'analyse doit reposer sur un examen des finalités et des missions de l'école, une localisation des enjeux réels, un travail d'élucidation des significations des choix opérés par un travail d'explicitation des arguments sur les plans politique, économique et pédagogique. Les plus récentes études comparatives internationales montrent que des efforts énormes ont été faits pour surmonter les limites inhérentes à ce type de travail et les lacunes de la recherche plutôt à grande échelle.

Tous les articles le montrent, nous avons besoin d'un dialogue entre les deux instances. Il reste à se mettre d'accord sur le type d'agenda possible. Il ne faut pas non plus oublier que l'éducation est avant tout un projet social, et concerne tous les citoyens.

Faouzia Kalali

Université de Rouen, Institut Universitaire de Formation des Maîtres
Unité Mixte de Recherche Sciences Techniques Education Formation,
Ecole Normale Supérieure de Cachan

Edgar W. Jenkins

School of Education, Université de Leeds, Royaume-Uni

Bibliographie

BAUDELOT C. & ESTABLET R. (2009), *L'élitisme républicain. L'école Française à l'épreuve des comparaisons internationales*, Paris, Seuil.

FENSHAM P. (2007), « Interest in science : lessons and non lessons from TIMSS and PISA », *Contribution from science education research*, R. Pinto & D. Couso (éd.), Dordrecht, Springer, p.3-10.

JENKINS E.W. (2000), « Research in Science Education : time for a Health Check ? », *Science Education*, vol.35, p.1-25.

JENKINS E.W. (2006), « Reforming school science : a commentary on selected reports and policy documents », *Science Education*, vol.45, n°1, p.65-92.

KALALI F. (2010), *L'enquête ROSE en France (Relevance Of Science Education) : analyse statistique des populations scolaires de Paris et de Créteil*, Rapport d'étude, en consultation sur le lien www.roseproject.no/network/countries/france/ROSE-Kalali.pdf

MARTINAND J.-L. (1989), « Comment discuter des finalités de l'enseignement des sciences ? », *Cahiers pédagogiques*, n°278, p.18-19.

NORMAND R. (2003), « L'influence des organisations internationales sur les politiques d'éducation », *Education et sociétés*, n°12, p.73-89.

OLSEN R.V. & LIE S. (2006), « Les évaluations internationales et la recherche en éducation : principaux objectifs et perspectives », *Revue française de pédagogie*, n°157, p.11-26.

OCDE (2009), *Top of the Class - High Performers in Science in PISA 2006*, Paris, OCDE Publications.
<http://www.oecd.org/pisa/pisaproducts/pisa2006/42645389.pdf>

UE (2004), Increasing Human resources for science and technology in Europe. Report : Europe needs more Scientists. http://ec.europa.eu/research/conferences/2004/sciprof/pdf/final_en.pdf