

## Conclusion

Réginald Burton and Christophe Dierendonck

Volume 34, Number 2, 2011

URI: <https://id.erudit.org/iderudit/1024851ar>

DOI: <https://doi.org/10.7202/1024851ar>

[See table of contents](#)

### Publisher(s)

ADMEE-Canada - Université Laval

### ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

### Cite this document

Burton, R. & Dierendonck, C. (2011). Conclusion. *Mesure et évaluation en éducation*, 34(2), 113–117. <https://doi.org/10.7202/1024851ar>

## CONCLUSION

---

### *De la conception à l'analyse critique*

En consacrant un numéro thématique aux enquêtes internationales, la revue *Mesure et évaluation en éducation* avait pour objectif de dresser un état des lieux des évaluations supranationales, d'en proposer une description technique – voire une analyse critique – et d'en anticiper les évolutions futures.

L'article de Nathalie Loye a planté le décor en offrant un panorama global de la douzaine de programmes d'enquête à large échelle actuellement conduits dans les différentes parties du globe. À partir d'une série de critères (objectifs, périodicité, dimensions évaluées, public-cible, mode de passation, format d'items, etc.), l'auteure a offert une comparaison des différents programmes d'évaluation. L'article de Dominique Lafontaine et de Gilles Raïche a décrit, dans un souci de vulgarisation scientifique, les principes techniques et méthodologiques des enquêtes internationales. L'accent a été mis sur les modèles de mesure utilisés et sur les procédures mises en place pour assurer la qualité et la comparabilité des données récoltées. Dans le troisième article, la problématique des enquêtes internationales a été abordée sous l'angle original de la sociologie politique. Sur la base d'entretiens et d'une analyse documentaire liée à PISA et au cas français, Xavier Pons a ainsi levé le voile sur les luttes d'influence politique et scientifique qui semblent se jouer lors de l'élaboration et de la mise en œuvre de la démarche d'évaluation au sein des différentes composantes institutionnelles de ce programme d'enquête. Enfin, dans l'article de Romain Martin et de Jean-Guy Blais, c'est une démarche prospective qui a été adoptée. Les auteurs se sont non seulement attelés à décrire les efforts actuels conduits en direction d'une évaluation, assistée par ordinateur, des compétences des élèves et des adultes, mais ils ont surtout attiré l'attention du lecteur sur les écueils et les défis majeurs de cette transformation progressive et visiblement planifiée des évaluations papier-crayon en évaluations par ordinateur.

Au terme de ce numéro spécial, il paraît opportun de dresser de manière synthétique les apports et les limites des enquêtes internationales évoquées tout au long de ce numéro spécial. Il semble en effet que l'intérêt notable de ces évaluations sur certains aspects soit partiellement contrebalancé par un lot de critiques formulées sur d'autres aspects.

Ainsi, le renouveau de l'éducation comparée doit sans aucun doute beaucoup à la multiplication et au développement des enquêtes internationales depuis la fin des années 1990. En proposant une prise d'information récurrente sur l'efficacité des systèmes éducatifs et donc, indirectement, sur la qualité de l'enseignement dispensé au travers d'une focalisation sur les résultats, les enquêtes internationales se sont progressivement installées au centre du débat éducatif. Les discussions menées autour des résultats du PISA 2000 ont par exemple été le déclencheur, dans plusieurs systèmes éducatifs, de profondes réformes scolaires ou d'action de régulation. Dans d'autres pays, elles ont fait office d'accélérateur des réformes en cours. Le caractère externe de ces évaluations semble également permettre une appréciation plus objective des compétences des individus et constitue, pour les responsables éducatifs, un outil de pilotage particulièrement intéressant puisque les enquêtes à large échelle permettent non seulement une comparaison entre systèmes mais également un suivi longitudinal. Les enquêtes à large échelle autorisent enfin l'étude des liens entre l'efficacité des systèmes éducatifs et leurs modalités d'organisation et de fonctionnement (systèmes à tronc commun *vs* systèmes à orientation précoce, phénomènes de ségrégation scolaire, degré d'équité, etc.). Elles visent à étudier le rendement scolaire au départ d'une approche multifactorielle et multiniveaux. Les bases de données extrêmement riches de ces enquêtes constituent d'ailleurs pour les chercheurs des sources uniques pouvant alimenter nombre d'analyses complémentaires nationales ou comparatives.

Les enquêtes internationales ne sont toutefois pas exemptes de reproches. Décrites comme des opérations de connaissance mais de plus en plus comme des opérations politiques, elles subissent des remises en question légitimes sur nombre d'aspects. On critique ainsi les valeurs qui les sous-tendent ou qu'elles véhiculent (libéralisme, utilitarisme, impérialisme, etc.). On leur reproche de ne pas prendre suffisamment en compte les caractéristiques des contextes nationaux, aboutissant à des conclusions erronées ou empêchant de la sorte tout transfert de politiques ou de pratiques efficaces d'un système éducatif à un autre. On cherche à les prendre en défaut sur des aspects de validité (postulats du modèle de mesure utilisé qui ne sont pas forcément assurés, dimensions évaluées restrictives, biais culturels, problèmes de traduction des items, etc.) et de méthodologie (démarche d'élaboration des items influencée politiquement, inégale participation des pays dans le processus d'élaboration des items, hermétisme du consortium international

face aux critiques et propositions méthodologiques de certains experts, remise en question du format des items, instabilité des construits mesurés au fil du temps, stratégies de contournement de certains pays lors de la constitution des échantillons, etc.). Enfin, le risque est élevé de retrouver des justifications ou des interprétations abusives des résultats des enquêtes internationales (cf. l'argumentaire des biais évoqué dans ce numéro par Xavier Pons ou les palmarès publiés dans les médias, par exemple), soit par souci de simplification ou de vulgarisation, soit par instrumentalisation politique.

### ***Des enquêtes internationales aux évaluations locales***

Le changement de paradigme qui s'est opéré depuis une vingtaine d'années en direction d'une gestion et d'un pilotage des systèmes éducatifs centrés sur les résultats (et non plus sur les inputs) a alimenté directement le phénomène de mondialisation des dispositifs d'évaluation qui concerne en premier lieu les enquêtes internationales. Le besoin d'évaluer la qualité des systèmes éducatifs au travers des résultats qu'ils engendrent en termes d'acquis scolaires a suscité le développement accru de plusieurs programmes internationaux bien connus tels que les études de l'IEA ou le PISA de l'OCDE. À quelques nuances près, ces programmes internationaux véhiculent une vision commune de l'évaluation tant sur le plan des contenus évalués (compétences en lecture, en mathématiques, en sciences et en résolution de problèmes) et des formes (test papier-crayon, QCM, questions ouverte à réponse courte, etc.) que des modèles de mesure utilisés pour le traitement des données (essentiellement les modèles de réponse à l'item).

Les conceptions internationales de l'évaluation, avec la considération croissante que certains acteurs du monde de l'éducation leur ont portée, se sont « imposées » sur le plan local. Ainsi, dans un mouvement global d'uniformisation, nombre de dispositifs d'évaluation locaux ou nationaux (épreuves d'orientation au Luxembourg, évaluations externes en Belgique, évaluations diagnostiques en France, etc.) ont été adaptés ou se sont inspirés des normes défendues dans le cadre des programmes internationaux.

La question se pose dès lors de savoir si cette transposition est légitime et souhaitable. Sur le plan méthodologique, est-il effectivement légitime d'utiliser dans des évaluations locales les modèles de mesure utilisés dans les enquêtes internationales ? Au plan idéologique, est-il souhaitable de focaliser les évaluations locales sur les contenus appréciés par les évaluations internationales ? Quelles en sont les conséquences sur l'enseignement ?

### *De la nécessité d'une transposition prudente*

La technicité des modèles de réponse à l'item (MRI), leur nature stochastique et leurs composantes statistiques relativement complexes n'ont pas toujours permis une approche cartésienne et prudente de leur exploitation. On a cru pendant longtemps que la transposition de ces modèles des enquêtes internationales aux évaluations locales était naturelle et ne souffrait d'aucun obstacle fondamental. D'un point de vue empirique, ce n'est que depuis l'avènement des micro-ordinateurs et surtout depuis l'évolution de la puissance des micro-processeurs, ces vingt dernières années, que nous disposons de la puissance de calcul nécessaire pour mettre à l'épreuve des échafaudages théoriques qui se sont parfois révélés bancals. C'est donc récemment que l'on commence à cerner empiriquement les limites des MRI et que les psychométriciens ont mis à la disposition des utilisateurs les résultats des premières simulations. Or, les objectifs, les caractéristiques et les exigences des évaluations locales sont parfois bien éloignés des préoccupations internationales. Nous nous interrogerons donc sur l'opportunité d'utiliser les MRI dans les dispositifs d'évaluation locales dont l'objectif est d'estimer de manière individuelle la compétence des sujets à des fins formative (pilotage des systèmes éducatifs), certificative (délivrance de diplôme) ou encore prédictive (orientation scolaire). Car, dans ces conditions, contrairement aux enquêtes internationales, la précision des estimations individuelles des sujets est capitale. Or, les études menées (Burton, à paraître<sup>1</sup>) montrent que les caractéristiques des évaluations locales en termes d'échantillons des sujets et d'items, de conditions d'application et de sélection des items ont une grande influence sur la précision de l'estimation de la compétence des sujets par les MRI. Les phénomènes observés dans certaines évaluations locales (multidimensionnalité, dépendance locale, fonctionnement différentiel des items, omissions dans les tests de vitesse) combinés aux caractéristiques des échantillons utilisés (effectif faible, distributions asymétriques) peuvent conduire à des imprécisions remettant en question le traitement équitable des enfants soumis à un dispositif d'évaluation dont les mécanismes sous-jacents et les enjeux les dépassent. Au demeurant, la technicité des méthodes et des modèles utilisés restent un sérieux obstacle à l'appropriation des résultats par les acteurs du monde de l'éducation qui éloigne les praticiens du processus d'évaluation.

L'influence des enquêtes internationales sur les évaluations locales ne concerne pas uniquement la méthodologie mais aussi les disciplines évaluées, les cadres de référence, les contenus et les formats d'items. Nous devons

prendre garde à ce que l'uniformisation des procédures d'évaluation ne masque les originalités de certains modèles d'évaluation diagnostique ou les particularités de certaines populations scolaires minoritaires. Les standards internationaux tirent leur légitimité d'un souci de comparabilité des résultats des évaluations. Sur le plan national, ce souci n'existe pas et d'autres objectifs sont visés (évaluations diagnostiques, pilotage des systèmes éducatifs, orientation, etc.). Il est dès lors souhaitable d'adapter les dispositifs d'évaluation locales non pas aux standards internationaux mais bien aux spécificités nationales.

**Réginald Burton**

**Christophe Dierendonck**

*Université du Luxembourg*

