

Mesure et évaluation en éducation



Retour sur l'évaluation diagnostique

Géry Marcoux and Sébastien Béland

Volume 39, Number 3, 2016

URI: <https://id.erudit.org/iderudit/1040138ar>

DOI: <https://doi.org/10.7202/1040138ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this document

Marcoux, G. & Béland, S. (2016). Retour sur l'évaluation diagnostique. *Mesure et évaluation en éducation*, 39(3), 97–103. <https://doi.org/10.7202/1040138ar>

Retour sur l'évaluation diagnostique

Géry Marcoux

Université de Genève

Sébastien Béland

Université de Montréal

Ce numéro spécial de *Mesure et évaluation en éducation* a pour ambition de présenter trois articles traitant de l'évaluation diagnostique dans le monde de l'éducation. Comme il a été précisé dans l'introduction et dans l'article de Rey, l'idée même du diagnostic en général et, plus spécifiquement, dans le monde scolaire n'est pas totalement neuve. Nous pourrions même avancer qu'elle est probablement concomitante à celle de l'instruction. En effet, sans en porter le nom, la maïeutique de Socrate (5^e siècle avant J.-C.) ne portait-elle pas déjà en soi les prémices du diagnostic? Quoi qu'il en soit, aujourd'hui, cette notion fait débat dans une partie du monde éducatif ou, du moins, interroge. De notre point de vue, au regard de l'utilisation expansive de ce concept au cours des dernières années (Eurydice, 2009), l'interrogation est légitime et les éclaircissements sont nécessaires (Marcoux, Fagnant, Loye & Ndinga, 2014).

Alors, que faut-il entendre par évaluation diagnostique en éducation et, plus singulièrement, dans le monde scolaire? Pour notre part, nous partirons de la proposition de Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman et Provus (1980), qui définissent l'évaluation comme un « processus par lequel on délimite, obtient et fournit des informations utiles permettant de juger des décisions possibles » (p. XXVI). Celle-ci a pour intérêt d'embrasser les nombreux développements de la notion d'évaluation des dernières décennies et, partant, de celle de l'évaluation diagnostique. Ainsi, nous pouvons reprendre et lire les définitions proposées en introduction, au regard de celle de Stufflebeam et ses collaborateurs (1980), qui est plus large. Dans ce cadre, l'évaluation diagnos-

tique a pour ambition de récolter des informations – « les paramètres significatifs » désignés par Legendre (2005, p. 397) – pour porter un jugement – par exemple, définir les forces et les faiblesses d'un sujet ou d'un groupe de sujets (Tatsuoka, 2009).

Ce choix nous permet aussi d'englober les différentes conceptions proposées dans les trois articles présents dans ce numéro et d'organiser leur analyse en proposant une grille de lecture affinée des trois niveaux : 1) la récolte d'informations, 2) le jugement, et 3) la prise de décision.

Dans ce cadre, l'article de Bernard Rey trouve toute sa place lorsque l'auteur interroge le concept d'évaluation diagnostique appliqué à la notion de « compétences scolaires », qui a émergé dans les années 1990. En effet, ce concept existait bien avant dans le champ scolaire francophone (de Landsheere, 1979, p. ex.) ; l'apparition des compétences et l'exigence faite au monde éducatif et aux maîtres des écoles obligatoires de les enseigner à leurs élèves ont renouvelé la question du « diagnostic ». La réflexion de l'auteur porte alors sur le statut, sur les conditions et sur les implications d'une évaluation diagnostique dans le champ scolaire sur la base d'exemples empruntés à des situations professionnelles.

Le second article est interrelié au premier, mais dans un contexte nord-américain et avec un paradigme de développement d'épreuves standardisées à visée diagnostique quelque peu différent. Nathalie Loye et Josée Lambert-Chan nous y présentent le développement documenté de l'outil diagnostique qu'elles ont conçu. Cet outil correspond pleinement aux attentes définies ci-dessus. De plus, le souci de transparence et de rigueur des auteures, dans la présentation de la démarche de construction de leur épreuve et dans la détermination des attributs qui la composent, fait preuve des qualités mais également des interrogations d'un travail scientifique exigeant dans ce domaine.

C'est un travail de même nature – en matière de rigueur et de transparence – que nous présentent Dominique Casanova et Marc Demeuse, dans le troisième article. Ici, l'évaluation diagnostique est élargie aux évaluateurs, et ce, afin de « mieux les accompagner dans leur pratique » et d'« améliorer la qualité du dispositif d'évaluation » (p. 61). Aussi, cherchant à garantir la validité des résultats d'épreuves complexes recourant à un double jugement humain, ces auteurs comparent deux méthodes : la

théorie classique des tests et le modèle de Rasch. Dans le cas qui les intéresse, il s'agit de pouvoir apparier les correcteurs selon leur profil diagnostique pour limiter au maximum les erreurs de mesure.

Bien que différents, ces textes cadrent bien avec le thème de l'évaluation diagnostique et sont, de notre point de vue, conciliables puisqu'ils offrent un spectre de possibles lorsque ce thème est évoqué. Au-delà, leurs points communs résident, pour une part, dans une même visée : un diagnostic pour améliorer l'évaluation. Ainsi, bien qu'ayant des approches différentes, les textes de Rey et de Loye et Lambert-Chan cherchent à affiner le diagnostic en vue de la régulation des apprentissages ; celui de Casanova et Demeuse vise surtout l'amélioration des appariements d'évaluateurs pour améliorer le jugement. De même, et pour une autre part, ces trois textes font preuve d'un souci pédagogique évident, en ce sens qu'ils ont chacun pour ambition de guider le lecteur pas à pas au cœur de ses réflexions scientifiques.

Pour le reste, ces trois textes peuvent être mis en dialogue et en débat. Si nous repartons des trois niveaux établis sur base de la définition de Stufflebeam et ses collègues (1980), nous pouvons nous attarder, dans un premier temps, sur le niveau de « récolte d'informations ». Plus précisément, nous pourrions nous intéresser aux questions : Qui récolte les informations ? Comment sont-elles récoltées ? Qu'est-ce qui est récolté ? D'abord, une distinction classique est effectuée entre évaluation interne et évaluation externe. Si les deux derniers articles peuvent être catalogués dans la seconde catégorie, le texte de Rey interroge les bienfaits de chacune d'elles. Dans le cas de l'évaluation interne, c'est l'enseignant qui effectue la récolte des données. L'avantage de cette condition est qu'il connaît l'individu qu'il interroge, ce qui rend les informations récoltées hautement pertinentes dans une visée de régulation des apprentissages. L'inconvénient, non abordé par Rey, concerne les biais à l'égard des personnes évaluées, qui ont été mis en évidence au cours des décennies par les recherches, entre autres en docimologie. Les évaluations externes ont évidemment les avantages et inconvénients inverses. Dans le cas de Loye et Lambert-Chan, le choix de l'externalité rejoint ce souci d'objectivation propre aux évaluations externes. Il en est de même pour Casanova et Demeuse.

Au-delà de l'intérêt ou de l'inconvénient pour l'évaluateur de connaître l'individu évalué, viennent en complément la forme d'évaluation proposée par cet évaluateur et le jugement porté sur les informations récoltées (deuxième entrée). Quel diagnostic les évaluations externes à grande échelle permettent-elles de faire? Permettent-elles de définir les forces et les faiblesses des apprenants et, partant, de donner des pistes aux enseignants? Quelle interprétation est faite par l'évaluateur sur la base des informations qu'il a récoltées? L'outil choisi détermine les informations récoltées, ce que démontre le texte de Loye et Lambert-Chan. Ce qui nous semble en débat, c'est la forme de ces évaluations standardisées. Pour Rey, le contexte même de la situation (p. ex., l'élève sait qu'il est occupé à passer un test dans un cadre scolaire déterminé pour une matière précisée), les formes d'énoncés proposés et les formats de réponses sollicités orientent l'action de l'individu et interfèrent avec la possibilité d'observer la compétence. Cette position est tenable si l'objet de l'évaluation est d'observer les compétences de l'individu. Toutefois, une autre position nous semble tenable. Si l'on considère que l'apprentissage de compétences nécessite la maîtrise et la mobilisation de ressources, on peut penser, comme le font Loye et Lambert-Chan, que d'observer les ressources maîtrisées par les élèves offre déjà de l'information, des pistes didactiques à travailler. C'est aussi oublier l'évolution des tests standardisés, qui sont passés de l'évaluation de la simple performance (sans doute limitée et normative) à des résultats plus détaillés (analyse de la démarche utilisée et intégration de nouvelles variables). Ces développements, tels qu'ils ont été présentés en introduction, ont fait passer, dans de nombreux cas, l'évaluation standardisée d'un paradigme psychométrique à un paradigme éducatif (voir Crahay, 2017 pour une proposition d'explication dans le monde francophone). Dès lors, l'apparition d'évaluations standardisées qui proposent des tâches nouvelles et complexes peut être envisagée (Dierendonck & Fagnant, 2014). De plus, l'utilisation des modèles de classification diagnostique, mentionnés dans l'article de Loye et Lambert-Chan, permet de raffiner considérablement les évaluations à grande échelle. Il est même possible de penser à des applications en testing adaptatif (Huebner, 2010), qui ouvrent la porte à des évaluations de tâches complexes assistées par ordinateur.

Le troisième et dernier niveau touche à la décision qui résulte du jugement. Il est alors important de relever, à l'instar de De Ketele (2010, p. 30), que la finalité du processus évaluatif n'est pas de «porter un jugement sur» mais de «fonder une prise de décision». Cette décision peut être de

différents ordres, mais sert souvent à établir, comme le souligne Allal (1979/1995), un plan d'action à des fins de régulation. La décision vise soit à assurer que les caractéristiques des élèves répondent aux exigences et contraintes du système, soit à faire correspondre les moyens de formation aux caractéristiques des apprenants (Allal, 2008). Or, d'autres dispositifs dits diagnostiques ont pour ambition d'anticiper le plus tôt possible l'apparition de difficultés d'apprentissage. La conception du diagnostic n'est alors plus de l'ordre de la régulation didactique de l'enseignant avec ses élèves, mais de celui du dépistage préalable à la mise en place de programmes d'aide aux apprentissages avant ou pendant l'école élémentaire. Ces évaluations visent à cibler les enfants présentant très tôt des difficultés d'apprentissage en vue d'effectuer des investigations complémentaires et de prendre des mesures d'aide et de prévention; de mettre ensuite en évidence pour ces enfants des signes cliniques permettant de suspecter la présence d'un trouble; et d'identifier précisément le trouble, la nature et l'ampleur des compétences déficitaires, les fonctions préservées et les différents facteurs associés (INSERM, 2007). L'idée et la problématique ne sont pas neuves! Toutefois, nous notons que, sous l'appellation diagnostique, pointe parfois une lecture médicale des difficultés d'apprentissage des élèves. Ainsi, sur la base d'une intention peut-être louable, qui consiste à dépister au plus tôt les troubles de l'apprentissage pour y remédier dans la foulée, émergent des instruments qui associent dans une démarche intégrative des professionnels de la santé, remisant l'enseignant à une tâche de repérage. Il y a donc nécessairement une vigilance à maintenir pour éviter de tomber dans la stigmatisation. Il est important, comme le souligne Rey, que «ce que l'évaluation fait apparaître, ce n'est pas un déficit qui serait intrinsèque à l'élève, mais un décalage entre sa manière d'interpréter la réalité et celle que l'école valorise» (p. 24). De plus, «l'élève ne peut plus être regardé sur le mode d'un mécanisme qui dysfonctionnerait et qu'il conviendrait de réparer; il ne peut pas non plus être appréhendé comme un malade qu'il faudrait guérir» (p. 24). Ce faisant, Rey renvoie à l'idée de la responsabilité de l'école, au risque d'oublier que l'école n'est pas un opérateur neutre, mais un opérateur transformant. En ce sens, l'échec de l'élève peut être aussi et avant tout l'échec de l'école (p. ex., en ce qui concerne le rapport au savoir). Par contre, cet échec peut aussi avoir un impact positif puisqu'il peut aider l'élève à maîtriser et intégrer les codes de l'école.

De ce point de vue, les trois articles cherchent des solutions, pour finalement réduire l'échec en travaillant tant sur les difficultés éprouvées par les auteurs que sur la qualité du travail des évaluateurs. Même si certaines tensions peuvent subsister entre les articles, nous les considérons comme complémentaires, car ils contribuent tous les trois à la recherche de solutions pour améliorer l'évaluation, ce qui constitue une problématique centrale en éducation.

RÉFÉRENCES

- Allal, L. (1979/1995). Stratégies d'évaluation formative: conceptions psychopédagogiques et modalités d'application. Dans L. Allal, J. Cardinet et P. Perrenoud (dir.), *L'évaluation formative dans un enseignement différencié* (6^e éd., pp. 153-183). Berne: Peter Lang.
- Allal, L. (2008). Évaluation des apprentissages. Dans A. van Zanten (dir.), *Dictionnaire de l'éducation* (pp. 311-314). Paris: Presses universitaires de France.
- Crahay, M. (2017). Introduction à l'axe 2: de la docimologie à la psychologie de l'évaluation. Dans P. Detroz, M. Crahay et A. Fagnant (dir.), *L'évaluation à la lumière des contextes et des disciplines* (pp. 133-140). Bruxelles: De Boeck.
- de Landsheere, G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris: Presses universitaires de France.
- De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, XV(1), 25-37. Récupéré de <http://www.cairn.info/revue-francaise-de-linguistique-appliquee-2010-1-page-25.htm>
- Dierendonck, C. et Fagnant, A. (2014). Approche par compétences et évaluation à large échelle: deux logiques incompatibles? *Mesure et évaluation en éducation*, 37(1), 43-82. doi: 10.7202/1034583ar
- Eurydice (2009). *Les évaluations standardisées des élèves en Europe: objectifs, organisation et utilisation des résultats*. Bruxelles: Eurydice.
- Huebner, Alan, (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation*, 15(3). Retrieved from <http://pareonline.net/getvn.asp?v=15&n=3>
- Institut national de la santé et de la recherche médicale (INSERM). (2007). *Expertise collective: dyslexie, dysorthographe, dyscalculie – Bilan des données scientifiques*. Paris: INSERM.
- Legendre, G. (2005). *Dictionnaire actuel de l'éducation*. Montréal: Guérin.
- Marcoux, G., Fagnant, A., Loye, N. et Ndinga, P. (2014). L'évaluation diagnostique des compétences à l'école obligatoire. Dans C. Dierendonck, E. Loarer et B. Rey (dir.), *L'évaluation des compétences en milieu scolaire et en milieu professionnel* (pp. 117-125). Bruxelles: De Boeck.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O. et Provus, M. M. (1980). *L'évaluation et la prise de décision en éducation*. Victoriaville, Canada: NHP.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York: Routledge/Taylor & Francis Group.