

L'évaluation en questions : plaidoyer pour une évaluation démocratique

Compte rendu de : Hadji, C. (2012). *Faut-il avoir peur de l'évaluation?* Bruxelles : De Boeck.

Par : Dany Laveault, Faculté d'éducation, Université d'Ottawa

La question que pose Charles Hadji dans le titre de son dernier ouvrage « Faut-il avoir peur de l'évaluation? » s'inscrit dans l'actualité récente. Tout comme son titre l'indique, l'ouvrage de Charles Hadji est une invitation au questionnement, philosophique sans doute, mais aussi citoyen, démocratique, méthodologique et éducatif. L'ouvrage s'inscrit dans la collection « Pédagogies en développement » de l'éditeur De Boeck. On y trouve de nombreux exemples tirés du monde de l'éducation, mais aussi des cas tirés de d'autres domaines (p.ex. finance, médecine, etc.) qui permettent de généraliser l'analyse de l'activité évaluative.

Le questionnement de l'auteur, auquel nous sommes conviés, se développe en trois temps. Dans la première partie, Hadji cherche à répondre à la question posée dans le titre en relevant des cas récents d'activités d'évaluation « calamiteuses ». L'étude approfondie de plusieurs cas types permet d'identifier de nombreuses calamités liées à l'activité d'évaluation elle-même. Se pose alors la question à savoir si l'évaluation peut être « sauvée ». Après avoir assumé que c'était le cas, l'auteur tente de répondre à une nouvelle question : « Que faut-il pour ne pas craindre l'évaluation? ». Suite à cette question, la deuxième partie débouche sur un nouveau questionnement autour de la méthodologie de l'évaluation. La troisième partie se conclue par la recherche d'une éthique de l'évaluation (partie 3).

Un questionnement ancré dans l'actualité récente

Dans la première partie, illustrée de nombreux cas tirés du monde de l'éducation, l'auteur n'a pas trop de difficultés à démontrer que l'évaluation puisse être une source de calamités. Il fait l'hypothèse, cependant, que le caractère « calamiteux » de l'évaluation ne provient pas de l'activité d'évaluation elle-même, mais de la société (p. 95). Ici, certains lecteurs pourraient ne pas être tout à fait du même avis et se poser les questions : « Quelle est la part de l'évaluateur dans cette calamité? » et « Qui est responsable des dommages de l'évaluation? ».

« On ne s'improvise pas évaluateur » (page 97) affirme Hadji. Plusieurs seront d'accord avec cette dernière affirmation même si elle ne résout pas le problème de la responsabilité de l'évaluateur en tant qu'acteur. Celui-ci est-il dans un « état agentique » comme dans la

célèbre expérience de Milgram sur l'obéissance? La co-production franco-suisse « Le jeu de la mort » (Nick, C., producteur, 2009) reprend justement cette tristement célèbre expérience pour démontrer le pouvoir terrifiant de la télévision. Nous pouvons y voir des personnes appelées à questionner un candidat lui administrer des chocs électriques de plus en plus douloureux à chaque nouvelle mauvaise réponse. Les évaluateurs sont-ils dans un « état agentique » comparable, du moins dans certaines situations?

Qu'il existe des standards encadrant les activités des concepteurs et des utilisateurs des résultats à des tests psychométriques (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2000; Ordre des conseillers et conseillères d'orientation et des psychoéducateurs et psychoéducatrices du Québec, 2003). constitue une bien piètre consolation. Lorsque Messick (1988) propose d'ajouter un nouveau critère de validité des résultats, celui de « validité de conséquence », plusieurs psychométriciens ne sont pas de cet avis, (voir notamment Reckase, 1998). Selon eux, cette exigence est impossible à satisfaire car ni le concepteur, ni l'utilisateur des résultats ne peuvent connaître à l'avance les effets qu'auront leurs tests. Mais les difficultés à estimer les risques associés à l'évaluation sont-elles des raisons suffisantes pour ne faire aucun effort en ce sens?

Un questionnement sur plusieurs plans

Après avoir remis l'évaluation en question et lui avoir accordé en quelque sorte un sauf conduit pour les deux parties suivantes, l'ouvrage traite d'une évaluation qui se questionne sur deux plans : le plan méthodologique (partie 2) et le plan éthique (partie 3). Ce double questionnement est essentiel : « Tout le travail d'évaluation repose bien d'abord sur l'explicitation d'un projet de questionnement » (page 191).

Au plan méthodologique, l'évaluation est légitime et « s'exerce dans les limites d'une saine raison » dans la mesure où elle s'inscrit « dans une démarche qui respecte, en prenant le temps qu'il faut pour cela, les trois moments méthodologiques de la spécification des attentes prioritaires, du choix judicieux des espaces de recueil d'informations, et de la quête des signes de satisfaction/insatisfaction des attentes dans chacun de ces espaces » (Hadj, page 196) . Sur ce point, l'auteur apporte peu de nouveau. Le courant behavioriste a largement contribué à établir ces trois moments quoiqu'en insistant probablement trop sur le caractère observable des informations recueillies et succombant à la trompeuse séduction du comptage et des chiffres.

Hadj a tout à fait raison de distinguer l'*évaluation estimative* qui « ne se traduit pas par un jugement de valeur, mais par un simple constat de fait » (page 164) de l'*évaluation appréciative*. Il faut bien dire que les anglo-saxons disposent d'un avantage linguistique sur les francophones. Ils peuvent avoir recours à deux termes différents pour distinguer

des moments précis de l'activité évaluative : « assessment » et « evaluation », le premier terme ayant été emprunté à la langue française et possède les mêmes racines que le verbe « assoir ». Le terme « assessment » se réfère donc à ce qu'Hadji décrit comme « des espaces judicieux de recueil d'informations » sur lesquels repose (« s'assoit »?) le jugement d'appréciation de l'évaluateur. Dans ce sens, « assessment » décrit toute démarche de collecte de preuves et de témoignages permettant de fonder ou de constituer les « assises » d'une évaluation appréciative, celle-ci pouvant être constituée d'une variété de données quantitatives (« mesure ») ou qualitatives (observations, témoignages, groupes de discussions, etc.).

Dans sa discussion sur les indicateurs, Hadji démontre bien que le recours à un seul indice, à des indices insuffisants de par leur nombre ou de par leur pertinence, peut conduire à des jugements faussés – inconsciemment ou pas. La constitution de la preuve (« assessment ») doit répondre à des critères méthodologiques rigoureux et –il en sera question dans la dernière partie – à des critères éthiques.

La « dictature de la moyenne » est sans doute le plus bel exemple des dangers qu'il y a à recourir à un seul indice statistique. Elle conduit les pédagogues à se centrer exclusivement sur la performance globale des groupes tout en gommant les différences individuelles. La variance est à cet égard un indice plus approprié pour déterminer l'équité des systèmes éducatif en déterminant si la progression se traduit par une diminution des écarts entre les élèves plutôt que par la progression d'un petit nombre, généralement parmi les meilleurs.

Plusieurs autres questions pertinentes sont abordées dans la deuxième partie, notamment toute la section portant sur la question de la modélisation de l'évaluation où l'auteur décrit les « fondamentaux de l'activité évaluative » qui constituent le premier impératif d'ordre méthodologique. Cette section donne l'occasion à l'auteur d'exposer et de discuter des modèles d'évaluation et de proposer ce qui serait en quelque sorte *un modèle d'évaluation des modèles d'évaluation*. Il accompagne ses discussions de nombreux exemples du monde de l'éducation et l'étude du cas de l'efficacité des ZEP en France constitue sans doute un bel exemple d'utilisation des indicateurs de *valeur ajoutée* (pages 142-143). Ceci étant dit, je demeure sceptique quant à l'utilisation de ces indicateurs qui, même s'ils permettent « d'échapper à des lectures hâtives, et à des interprétations calamiteuses, de simples résultats bruts » (page 41) laissent encore beaucoup de place à la manipulation de données (valeur ajoutée par rapport à quoi?) et pour lesquels les preuves sont encore à faire quant à leur impact sur l'évaluation « formative » des établissements (Laveault, 2009).

Une des contributions importantes du chapitre 2 porte sur une évaluation de l'*éducativité* : « l'évaluation ne prend tout son sens que si elle s'interroge sur la réalisation de ce qui donne sens à la pratique qu'elle évalue » (p. 157). Hadji ouvre ici

une perspective constructive aux évaluateurs soucieux d'estimer la validité de conséquence de leur activité. Il ne s'agit plus d'éviter que les résultats aient des conséquences négatives sur les examinés, mais bien de se demander : « ce qui a été fait représente-t-il une bonne solution pour le problème d'avoir à éduquer? » (p. 157).

Une évaluation « démocratique »

La partie 3 introduit le concept d'*évaluation démocratique* : « une pratique évaluative qui se mettrait vraiment au service de tous » (page 216). Pour y parvenir, l'évaluation démocratique privilégie les questionnements les plus utiles au plus grand nombre, les objets les plus importants pour les publics évalués et enfin les fonctions les plus utiles aux publics concernés. Ces choix démocratiques de l'évaluation se fondent sur l'éthique suivante : « contribuer au respect et à la sauvegarde de la dignité de l'être humain » (p. 282).

L'évaluation démocratique est-elle une utopie philosophique? En effet, comment améliorer l'évaluation, sans améliorer les évaluateurs? Parmi les questionnements utiles, il est peu question dans cette troisième partie de l'autoquestionnement de l'évaluateur, non pas dans une problématique de développement de l'évalué (p. ex. l'élève), mais dans la perspective du développement professionnel de l'évaluateur lui-même (p. ex. l'enseignant). La relation évaluative est constructive non seulement pour l'évalué mais aussi pour l'évaluateur. Toute la problématique des régulations suscitées par l'évaluation dans l'ajustement pédagogique mériterait qu'on s'y attarde davantage (p. 232), car l'évaluation de l'élève, utilisée à son plein potentiel, peut être source de perfectionnement professionnel pour l'enseignant la vie durant.

En réalisant les risques associés à l'évaluation et en prenant conscience des exigences méthodologiques et éthiques qui doivent gouverner l'activité évaluative, on pourrait se demander si la formation initiale et le développement professionnel des enseignants est à la hauteur de ces exigences qui feraient que l'évaluation deviendrait « libératrice » autant des capacités de l'enseignant que de celles de l'élève.

En définissant l'« évaluation formatrice », Nunziati (1990) insistait déjà sur le fait que l'autoévaluation était une faculté naturelle présente chez tous et que celle-ci avait besoin d'être raffinée. Les évolutions récentes dans le domaine de l'évaluation vont dans ce sens. Earl (2003) traite d'évaluation en tant qu'apprentissage (*assessment as learning*) et élève l'autoévaluation au rang de compétence à développer. Pour Earl, il ne suffit pas de faire de l'élève un collaborateur du processus d'évaluation, mais un acteur critique qui fait le lien entre évaluation et apprentissage. Le développement de l'évaluation en tant qu'apprentissage va tout à fait dans le sens de ce que Hadji (1997) proposait déjà comme principe majeur d'une éthique de l'*agir évaluationnel*: « N'exercer son pouvoir

d'évaluateur que s'il contribue à la prise de pouvoir sur soi de l'évalué ». Ce pouvoir de l'évaluateur ne peut contribuer à l'actualisation des capacités de l'élève que s'il se fonde sur de réelles compétences de l'évaluateur et une éthique de service.

L'évaluation formative des apprentissages des élèves s'est surtout concentrée sur la rétroaction à donner à l'élève et beaucoup moins sur la rétroaction inverse (Black & Wiliam, 2009, p. 24), soit le feed-back en retour de l'élève à l'enseignant. Les avancées récentes dans le domaine de l'évaluation vont tout à fait dans le sens du projet d'une évaluation plus démocratique et relèvent, tout comme le projet de Hadji, d'une logique développementale et humaniste des capacités humaines : « La seule chose qui légitime la position haute occupée de fait par l'évaluateur est l'espoir de permettre à l'évalué d'arriver à son tour en position haute : celle du sujet connaissant, compétent, maître de soi-même. [...] Le bon évaluateur est celui qui, à la fin, s'efface et disparaît » (p. 280).

L'évaluation peut être une source de déséquilibre, tant pour l'évaluateur (« Pourquoi tel élève ne réussit-il toujours pas? ») que pour l'évalué (« Pourquoi est-ce je n'y arrive pas? »). Pour que l'évaluation soit constructive, ce déséquilibre doit s'inscrire dans une équilibration « majorante » (Piaget, 1975) qui donne lieu à un dépassement de chacun, au niveau de l'apprentissage pour l'élève et du perfectionnement professionnel pour l'enseignant. Cette équilibration majorante requiert que soient établis de nouveaux équilibres, entre feedback direct et feedback inverse, entre pouvoirs de l'évaluateur et de l'évalué et capacités à évaluer, tant de l'évaluateur que de l'évalué.

Les réflexions récentes dans le domaine vont dans le sens d'une évaluation qui est davantage un dialogue, une rencontre entre l'évaluateur et l'évalué : « l'évaluation-soutien d'apprentissage fait partie des pratiques quotidiennes des élèves et des enseignants qui, individuellement et en interaction, réfléchissent sur et réagissent à l'information provenant d'échanges, démonstrations et observations afin de favoriser les apprentissages en cours » (Allal et Laveault, 2009, p. 102). Mais cette rencontre est aussi une rencontre de chacun avec soi-même, car nous sommes tous évaluateurs et évalués et que le dépassement de soi passe par une prise de conscience accrue de notre jugement, de ce qui le fonde et des conséquences qu'il a pour nous et pour les autres. C'est dans une perspective de dépassement de soi-même qu'une autoévaluation critique de l'évaluateur peut devenir un instrument de développement professionnel continu et d'autorégulation de sa double nature, d'évaluateur et d'évalué.

À la question posée par le titre de Hadji, il faudrait encore ajouter celle-ci : « Faut-il craindre de s'autoévaluer? ». Mais Hadji a déjà répondu à cette question en insistant sur l'importance du questionnement, méthodologique et éthique de l'évaluateur. En paraphrasant Socrate, nous pourrions répondre : « Évaluateur, évalue-toi toi-même ».

Références

- Allal, L. & Laveault, D. (2009). Évaluation-soutien d'apprentissage. Prise de position formulée par la Troisième Conférence internationale sur l'Évaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation*, 32(2), 99-107.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2000). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Black, P. & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Earl, L. (2003). *Assessment as learning. Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Hadji, C. (1997). Pour une éthique de l'agir évaluationnel. *Mesure et évaluation en éducation*, 20(2), 7-26.
- Laveault, D. (2009). L'amélioration de l'efficacité du système éducatif : sur quels indicateurs s'appuyer? In X. Dumay & V. Dupriez (Eds). *L'efficacité dans l'enseignement : promesses et zones d'ombre*. Bruxelles : De Boeck (p. 177-194).
- Messick, S. (1988) The once and the future issues of validity : Assessing the meaning and consequences of measurement . In H. Wainer & H.I. Braun, *Test Validity*. Hillsdale, NJ : Lawrence Erlbaum.
- Nick, C. (producteur). (2009). *Le jeu de la mort* [Documentaire télévisé]. Coproduction France Télévisions et Radio télévision Suisse.
- Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques*, 280, 48-64.
- Ordre des conseillers et conseillères d'orientation et des psychoéducateurs et psychoéducatrices du Québec (2003). *Normes de pratique du testing en psychologie et en éducation*. Montréal: Institut de Recherches Psychologiques.
- Piaget, J. (1975). *L'équilibration des structures cognitives. Problème central du développement*. PUF : Études d'épistémologie génétique, vol. XXXIII.
- Reckase, M.D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, 17(2), 13-16.