

Sommaire

Modèles et modélisations face à une crise du sens de l'évaluation en éducation <i>Lucie Mottier Lopez* & Gérard Figari**</i> <i>*Université de Genève **Université de Grenoble</i>	7
LES ACTIVITÉS D'ÉVALUATION : QUELS SAVOIRS PRODUITS ?	25
La production de savoirs négociés dans deux recherches collaboratives sur l'évaluation formative <i>Joëlle Morrisette*, Lucie Mottier Lopez** & Walther Tessaro**</i> <i>*Université de Montréal **Université de Genève</i>	27
Entre savoirs scientifiques et usages normatifs : les enquêtes internationales d'évaluation des acquis des élèves <i>Dominique Lafontaine & Christian Monseur</i> <i>Université de Liège</i>	45
Les évaluateurs, des épistémologues en actes ? <i>Claire Tourmen & Patrick Mayen</i> <i>AgroSup Dijon</i>	63
LES ACTIVITÉS D'ÉVALUATION : QUELLES MODÉLISATIONS MISES EN ŒUVRE ?	79
L'évaluation des élèves. Une modélisation interactionniste des pratiques professorales <i>Pierre Merle</i> <i>IUFM de Bretagne et Université européenne de Bretagne</i>	81
Élaboration et validation d'un questionnaire d'attitudes en éducation : considérations épistémologiques et métrologiques <i>Pierre Valois, Claude Houssemand & Astrid de Leeuw</i> <i>Université du Luxembourg</i>	97

Autorégulation et évaluation-soutien d'apprentissage <i>Dany Laveault</i> <i>Université d'Ottawa</i>	115
Les modes de pensée, organisateurs des modèles d'évaluation : pour la reconnaissance d'une épistémologie profane <i>Michel Vial</i> <i>Université de Provence</i>	131
Modélisation de la participation des acteurs dans le processus d'évaluation de la formation <i>Pedro Rodrigues* & Eusébio André Machado**</i> <i>*Université de Lisbonne **Université Portucalense</i>	147
D'UNE ÉPISTÉMOLOGIE ABSENTE À UNE ÉPISTÉMOLOGIE EN DÉBAT	163
Enjeux épistémologiques, et donc aussi politiques, de la méta-évaluation <i>Lise Demailly</i> <i>Université de Lille 1</i>	165
Les exigences inconciliables des activités évaluatives et leurs paradigmes de référence <i>Linda Allal</i> <i>Université de Genève</i>	181
À la recherche de paradigmes unificateurs ou de paradigmes intégrateurs ? <i>Jean-Marie De Ketele</i> <i>Université catholique de Louvain</i>	195
Configurations pour une épistémologie de l'évaluation <i>Gérard Figari</i> <i>Université de Grenoble</i>	211
Notices biographiques	225
Table des matières	233

Modèles et modélisations face à une crise du sens de l'évaluation en éducation

Lucie Mottier Lopez* & Gérard Figari**

***Université de Genève**

****Université de Grenoble**

Une abondante littérature existe sur l'évaluation en éducation depuis que celle-ci a envahi tous les niveaux et contextes des systèmes éducatifs et de formation : des comptes rendus de recherches, des analyses de pratiques, des rapports d'audits, des actes de colloques, des guides méthodologiques, des bilans et synthèses des travaux réalisés. Cette production s'accompagne de remises en question régulières à propos du *sens* même de la démarche évaluative. Des voix se font entendre pour critiquer la validité, la crédibilité, l'usage social des résultats produits par les pratiques évaluatives. Ce problème de sens qui affecte l'évaluation est parfois attribué au contexte plus général de crise des systèmes éducatifs et de formation, dont l'efficacité et l'équité ainsi que les modes de pilotage sont mis en cause. Selon certains, l'injonction toujours plus pressante au développement de l'évaluation à tous les niveaux du système éducatif contribue à alimenter cette crise (Demailly, 2006).

L'hypothèse que cet ouvrage propose d'explorer considère que ce sont les modèles mêmes de l'évaluation qui sont à interroger. En effet, le sens et la légitimité des résultats de l'évaluation sont liés aux modèles d'intelligibilité qu'elle convoque et à leur modélisation en tant que construction de systèmes symboliques et représentationnels par lesquels elle se définit et s'opérationnalise dans un contexte social et historique donné. Cette hypothèse ouvre le débat sur la nature intrinsèque des savoirs produits par l'évaluation au regard des démarches mises en œuvre, de ses finalités et

de ses idéologies. Elle pose la question d'une épistémologie possible, voire spécifique, aux évaluations en éducation. L'enjeu est alors d'essayer de dépasser une certaine *doxa* « évaluative », faisant implicitement autorité dans les systèmes et milieux éducatifs, pour discuter d'une possible *épistémè* de l'évaluation.

Même si la notion d'épistémologie est de plus en plus associée à celle d'évaluation (aussi bien dans les pratiques que dans les recherches), il n'existe pas, aujourd'hui, à notre connaissance, d'ouvrage spécialement dédié à la téméraire thématique de « l'épistémologie de l'évaluation ». Trois raisons, au moins, peuvent l'expliquer :

1. Lorsque l'évaluation est abandonnée à ce que Fournet (2008) appelle un « fonds émietté, réifié et disparate dont la pratique ne traduit que des cristallisations partielles », rares sont ceux qui songeraient encore à lui chercher une signification singulière présentant un lien quelconque avec un objet de savoir (à transmettre, à expérimenter, à construire).
2. Les travaux se préoccupant d'évaluation, inscrits dans des disciplines comme la sociologie, la psychologie cognitive et la psychologie sociale, par exemple, produisant leurs propres travaux sur ce thème, construisent des questionnements épistémologiques en rapport avec leur propre discipline.
3. Les chercheurs et les praticiens occupés par l'évaluation en éducation remettent peu en question la distinction entre pratique, d'une part, et recherche sur la pratique, d'autre part. Des postures se démarquent, certes, dans la littérature ; Bedin (2008), par exemple, établit une liaison entre deux dimensions indissociables, selon elle, de l'évaluation : « l'acte social » et « l'objet de recherche », autrement dit, la « praxis de l'évaluation » et la « praxéologie de l'évaluation », l'une ne fonctionnant pas sans l'autre. Mais dans ce cas, il s'agit surtout de postures de questionnement.

À partir d'interrogations critiques sur quelques résultats produits par l'évaluation dans différents contextes de recherche et de pratique, cette introduction visera à expliciter les enjeux scientifiques de la problématisation qui est au cœur de cet ouvrage : questionner les modèles et modélisations de l'évaluation en éducation dans une perspective épistémologique. Plusieurs paradigmes et modèles de l'évaluation disponibles dans la littérature seront exposés afin, déjà, de souligner la diversité de leurs conceptualisations et usages. Au regard de cette diversité, les enjeux d'une réflexion critique sur les modélisations pour l'évaluation seront énoncés afin d'ouvrir le débat qui traverse l'ensemble des contributions de cet ouvrage autour de cette question : dans quelle mesure une possible reconstruction épistémologique permettrait-elle de penser une *épistémè* propre à l'évaluation qui permettrait à son tour, peut-être, de dépasser un certain éclatement de ces paradigmes et modèles ? Enfin, la présentation de l'organisation générale de l'ouvrage et des chapitres qui le composent viendra clore cette introduction.

PROBLÉMATIQUE DE L'ÉVALUATION COMME CRISE DU SENS

Trois exemples sont discutés ci-après dans le but d'initier le débat épistémologique qui nous intéresse. Ces exemples n'ont pas valeur de catégories, mais ils montrent trois façons complémentaires de poser le problème.

Une mise en doute du jugement évaluatif des enseignants

Le premier exemple concerne les études bien connues sur le « jugement évaluatif » de l'enseignant. Dès la fin du XIX^e siècle en Europe et aux États-Unis, des travaux ont remis en question la fiabilité des pratiques de notation aux examens, en démontrant notamment les disparités de jugement entre évaluateurs (par exemple Bovet, 1914 ; Edgeworth, 1888 ; Starch, 1917, cités par Piéron, 1963). Les études docimologiques de Piéron (1963), de de Landsheere (1971), de Bonboir (1972), entre autres, produisent comme résultats la mise en évidence des *biais* qui affectent les jugements scolaires (« effets » d'ordre, de contraste, d'assimilation, de contamination, de halo, infidélité dans le temps, etc.¹). Ces biais sont justifiés par le recours au « paradigme de la mesure » et à ses critères de validité pour démontrer les « défaillances » du jugement scolaire. Les postulats et modèles défendus par le paradigme de la mesure sont ici considérés comme un cadre légitime de référence pour conférer la rationalité manquante aux jugements intuitifs et subjectifs des enseignants, en cohérence avec le mouvement positiviste qui a marqué, entre autres, une recherche d'assise scientifique aux sciences humaines, dont font partie les sciences de l'éducation.

D'autres travaux privilégient, quant à eux, une approche interprétative pour comprendre les phénomènes qui sous-tendent les jugements d'évaluation tels qu'ils s'exercent dans des pratiques socioculturelles et institutionnelles. La conception sous-jacente ici est que « l'appréciation faite par les enseignants dans le cadre des activités quotidiennes de la salle de classe est un processus qui relève plus de l'ethnographie ou du processus clinique que de la mesure » (Blais, 2005, p. 125). Dans cette perspective, ce qui est désigné comme un biais dans une approche docimologique n'est plus automatiquement considéré comme une défaillance du jugement scolaire. Il est injuste, sans conteste, que l'ordre des copies ait une influence sur la note attribuée, tout comme certaines caractéristiques non scolaires de l'élève. Par contre, la prise en compte dans le jugement professionnel de l'enseignant de certains facteurs, tels l'investissement de l'élève dans le travail scolaire, ses projets, l'appui familial, etc., est considérée comme pertinente pour soutenir la régulation des progressions

1. Voir par exemple Amigues et Zerbato-Poudou (1996) pour une définition et une discussion critique de ces principaux biais (appelés « effets » dans la littérature évoquée).

des élèves (Mottier Lopez & Allal, 2008). Si d'un point de vue métrique ces facteurs font effectivement varier le jugement de l'enseignant, ils prennent sens par rapport à la finalité de la pratique éducative.

Dans le cas rapporté, l'approche docimologique nous alerte sur les variations des jugements scolaires mais sans toutefois donner une explication totalement satisfaisante. Nonobstant les propositions d'amélioration suggérées par les docimologues, les constats en effet persistent : le jugement scolaire de l'enseignant reste sous l'influence de certains facteurs (voir Bressoux & Pansu, 2003). L'approche interprétative située apporte un autre éclairage en tentant d'appréhender la logique écologique des pratiques évaluatives des acteurs. Les critères de validité traditionnellement appliqués aux évaluations scolaires à partir de ceux définis par le paradigme de la mesure (pertinence, validité, fiabilité, fidélité) sont alors soumis à débat. D'autres critères sont élaborés par les chercheurs, par exemple à partir des propositions des approches qualitatives et de leur épistémologie (voir Allal, 2009 ; Guba & Lincoln, 1989 ; Mucchielli, 1996).

Ce premier exemple vise à rappeler, si cela était nécessaire, que les résultats des recherches varient en fonction des choix théoriques dans lesquels ils s'ancrent. Une hypothèse est que la diversité et cette relativisation du sens en fonction de l'ancrage épistémologique ont pour conséquence de contribuer à un éclatement des explications possibles et à un « brouillage » du champ de l'évaluation.

Une évaluation contestée de l'« efficacité » des écoles

Le deuxième exemple porte sur une pratique d'évaluation des écoles critiquée par Blais *et al.* (1999). Sans remettre en question l'importance pour tout système éducatif de s'évaluer et d'être évalué, les auteurs effectuent une analyse critique d'une étude commanditée par l'Institut économique de Montréal s'intitulant *Bulletin des écoles secondaires du Québec* (Cowley & Marceau, 2000). Dans cette étude, le bulletin scolaire fonctionne comme une mesure indépendante de la performance annuelle de 474 écoles secondaires. Le but est d'utiliser les résultats des élèves à des épreuves standardisées du ministère de l'Éducation pour établir un classement des écoles. Blais *et al.* contestent les résultats de cette étude évaluative en critiquant les points suivants :

- L'idéologie sous-jacente qui guide le choix des indicateurs de l'évaluation des écoles, ancrée dans une politique de libre marché et dont rien ne garantit qu'elle concourt à rehausser réellement la qualité du système éducatif ;
- La problématisation de l'objet évalué et sa conceptualisation : ici, le rendement d'une école qui se mesure par les résultats scolaires des élèves ;

- La méthode choisie pour appréhender la qualité des écoles : choix des données, variables retenues, construction des indicateurs, procédés métriques et statistiques utilisés ;
- Les stratégies de diffusion des résultats de l'étude, suspectées de chercher à influencer l'opinion publique (notamment les parents des élèves) en faveur d'un libre marché en éducation et de la privatisation du système éducatif au Québec.

Cet exemple souligne la spécificité du raisonnement évaluatif qui réside dans la possibilité de conceptualiser trois dimensions interreliées : 1) une théorisation de l'objet évalué ; 2) une théorisation du fondement méthodologique de l'évaluation de l'objet ; 3) une théorisation des finalités et de l'exploitation des résultats de l'évaluation. La conceptualisation de ces dimensions relève d'un processus de modélisation qui demande à construire les faits à évaluer (considérant que la réalité n'est pas directement appréhendable), à déterminer les méthodes pour évaluer les faits construits, et à expliciter l'usage social de l'évaluation (ou ses fonctions²) – considérant que cette dernière, en tant qu'activité finalisée, située dans un contexte social, historique et politique donné, n'est jamais neutre.

Une critique des évaluations dans l'enseignement supérieur

Le troisième exemple s'inscrit dans le contexte de l'enseignement supérieur dans lequel l'évaluation formative, qui vise à améliorer les pratiques d'enseignement universitaire et renforcer les apprentissages des étudiants, est distinguée de l'évaluation à visée administrative conçue dans une logique de pilotage institutionnel et de gestion des ressources humaines et des carrières académiques. Toutefois si ces fonctions de l'évaluation sont différenciées au plan conceptuel, « la réalité des pratiques montre qu'il n'est pas si aisé de distinguer les deux approches et qu'il faut renoncer à l'illusion d'une séparation claire et distincte » (Rege Colet, 2009, p. 282). Le compte rendu de Pelletier (2009), expert ayant participé à de nombreuses démarches d'évaluation d'établissements ou de programmes de l'enseignement supérieur dans plusieurs pays, le confirme : les approches sont combinées et présentent un caractère foncièrement ambigu pour les acteurs concernés qui développent alors des stratégies diverses en fonction de leurs statuts (étudiants, enseignants, responsables de programmes, dirigeants) et de leur rapport à l'évaluation³. Quel est alors le sens des résultats produits

2. Considérant qu'il y a des fonctions préexistantes, des fonctions effectives, des fonctions cachées, des fonctions détournées.

3. L'étude de Nagels et Vourch (2009) dégage, parmi les différents acteurs concernés par l'évaluation des enseignements et des programmes universitaires, trois types de

par l'évaluation, à quoi servent-ils ? Pelletier est plutôt catégorique : « Si l'on convient qu'évaluer c'est porter un jugement, juger est une activité "politique" et il n'y a pas de "jugement" qui peut être scientifique... évaluer relève de l'art et de la technique, de l'activité politique et, à la limite, de la réflexion philosophique » (2009, p. 106). De ce point de vue, l'évaluation produit essentiellement des savoirs locaux stratégiques et relatifs. Saisie dans un ensemble complexe de jeux sociaux d'influences et de pouvoir, l'évaluation semble ici perdre sa fonction de mise à distance des faits évalués et de construction collective de sens.

Cet exemple est significatif d'une certaine doxa qui consiste, entre autres, à réduire l'évaluation à un « jugement de valeur ». Cette doxa tend à affirmer alors que l'évaluation n'est pas en mesure de produire des interprétations fiables des données recueillies à travers une démarche évaluative, de comparer les résultats à d'autres résultats, de les trianguler à diverses sources d'informations objectives et subjectives, de les contester si nécessaire. Si on s'arrête à cette doxa, on réfute l'idée que l'évaluation peut contribuer, avec d'autres démarches, à l'explication et à la compréhension des phénomènes. L'évaluation ne serait-elle vraiment qu'une accumulation de jugements personnels, de réflexions politiques, de considérations philosophiques pouvant faire l'objet d'interprétations multiples opposables les unes aux autres ? Affirmer que l'évaluation n'est qu'un « jugement de valeur », n'est-ce pas une façon de renoncer à l'objectif de dégager à travers la démarche évaluative un ensemble d'énoncés cohérents, reconnus et validés par une communauté scientifique ou professionnelle ?

L'évaluation désigne de fait une « réalité » plurielle, entre interprétations d'informations situées et recherche. Comme nous l'avons montré avec les trois exemples exposés, l'ancrage théorique et méthodologique adopté par les chercheurs conduit à la production d'un sens particulier du phénomène étudié. D'où l'importance, pour comprendre les résultats produits par les démarches évaluatives et la recherche en évaluation, d'en expliciter l'ancrage.

DIVERSITÉ DES CONCEPTUALISATIONS DE L'ÉVALUATION EN ÉDUCATION : LES ENJEUX DE LA MODÉLISATION

Dans la littérature, les chercheurs utilisent des terminologies variées pour conceptualiser l'évaluation en éducation. Dans cette introduction, nous commençons par distinguer trois grandes catégories : les paradigmes de l'évaluation, les modèles d'évaluation, les méthodes d'évaluation. Le

représentation : 1) l'évaluation est jugée illégitime ; 2) elle est perçue uniquement comme un dispositif de sanction ; 3) elle a une valeur pragmatique à des fins d'amélioration et de pilotage.

chapitre conclusif de Gérard Figari proposera une réorganisation conceptuelle face à un contexte général de « pensée éclatée » dont nous faisons brièvement état ci-après.

Les paradigmes de l'évaluation

Le paradigme, au sens de Kuhn (1970), désigne un ensemble cohérent de postulats regroupant « les lois, des théories, des applications et des dispositifs expérimentaux qui doit fournir des modèles qui donnent naissance à des traditions particulières de recherche » (p. 30). Plusieurs auteurs se sont emparés de cette notion pour envisager des *paradigmes de l'évaluation*. C'est le cas de Rodrigues (2006) lorsqu'il définit trois grandes « perspectives » d'évaluation à partir d'une analyse de nombreuses publications, notamment anglo-saxonnes : 1) la perspective objectiviste ou technique, caractérisée par l'évaluation externe, la causalité linéaire, la méthodologie expérimentale, la comparaison ; 2) la perspective subjectiviste ou pratique, caractérisée par l'autoévaluation et l'observation participante ; 3) la perspective dialectique et critique, caractérisée par la coévaluation et la causalité circulaire. Ces trois perspectives se rapprochent des grandes « conceptions de l'évaluation »⁴ définies par Cardinet dans sa contribution de 1990, à savoir : 1) l'évaluation externe et la gestion systémique, associée au modèle des sciences de la nature ; 2) l'évaluation interne et la gestion autonome, associée au modèle des sciences humaines ; 3) l'évaluation négociée et la gestion interactive, associée au modèle des sciences compréhensives. Le projet des deux auteurs est de circonscrire les logiques de l'évaluation au regard de leur rapport à la production de connaissances dans différentes perspectives scientifiques. Les conceptualisations de l'évaluation se situent alors explicitement sur un plan épistémologique, mettant en avant l'impact des paradigmes de l'évaluation sur les postulats théoriques défendus, sur les choix méthodologiques, sur l'implication des acteurs, sur le type de savoirs produits par la démarche évaluative.

Dans la littérature, la notion de paradigme de l'évaluation n'a pas seulement été exploitée dans la lignée de la définition de Kuhn (1970). Dans sa contribution de 1993, De Ketele énumère les courants de pensée qui, selon lui, ont eu un impact sur les pratiques d'évaluation en éducation et leur conceptualisation. L'auteur distingue 10 « paradigmes de l'évaluation », parmi lesquels le paradigme de l'intuition pragmatique, le paradigme docimologique, le paradigme sociologique, le paradigme de l'évaluation centrée sur les objectifs, le paradigme de l'évaluation formative dans un enseignement différencié, le paradigme de l'évaluation

4. Dans son texte, Cardinet parle aussi de « modèles » ou d'« approches » pour désigner les trois conceptions dégagées et les sciences (de la nature, humaines, compréhensives) auxquelles il les associe.

au service de la décision, etc. La définition de ces « paradigmes » se situe sur différents plans, certains relevant d'un ancrage explicitement disciplinaire, d'autres de courants pédagogiques, voire de méthodes ou de principes d'action. Ici, le paradigme de l'évaluation ne se définit pas forcément au regard de son rapport à la science et aux lois qui la définissent, comme c'est le cas par contre dans les contributions de Cardinet et de Rodrigues citées plus haut. De Ketele (2001) analyse l'évolution des paradigmes de l'évaluation tels qu'il les avait identifiés en 1993. Il constate des pertes de vitesse pour certains mais sans qu'ils disparaissent totalement, des enrichissements pour d'autres par un processus d'élargissements successifs. L'auteur s'interroge : « On peut se demander si un nouveau paradigme de nature plus englobante n'est pas en train de se construire progressivement » (2001, p. 107). Dans cette contribution, De Ketele poursuit la réflexion en interrogeant la pertinence d'un paradigme possiblement unificateur et intégrateur.

Les modèles d'évaluation

D'une façon générale, la notion de modèle est appliquée à nombre de contextes et pour des utilisations très diverses : modèle théorique, modèle d'analyse, modèle mathématique, modèle technologique, modèle professionnel, etc. Dans cet ouvrage, nous nous intéressons plus spécialement au modèle en tant que représentation conceptuelle qui permet de rendre intelligible une situation, un phénomène ou un processus (Le Moigne, 1999), à des fins scientifiques mais pouvant également être orienté vers l'action en situation. Mouloud (2002) définit le modèle en sciences sociales et humaines comme « une figuration et en même temps un schéma directeur [...] une transcription abstraite mais contrôlée par la pensée logique d'une réalité concrète, empirique, dont l'étude directe ne donnerait que des relations approximatives » (p. 294). La notion de modèle suppose une mise en relation entre une « théorisation » et la concrétude de la réalité empirique expérimentée.

Comme on l'a observé pour la notion de paradigme, la littérature scientifique s'est emparée de l'idée de « modèles de l'évaluation ». On y trouve à nouveau des niveaux de conceptualisation variés. Ainsi, certaines propositions relèvent de modèles généraux, explicitement disciplinaires, comme c'est le cas du « modèle sociologique de l'évaluation scolaire » (voir Demailly, 2006 ; Merle, 1998). D'autres apparaissent plus spécifiques afin de circonscrire, par exemple, certaines pratiques pédagogiques, tel le « modèle de l'évaluation formative dans un enseignement différencié » (voir Allal, Cardinet & Perrenoud, 1979). Quant aux propositions de Bonniol et Vial (1997), elles déterminent trois grands modèles (l'évaluation-mesure, l'évaluation-gestion, l'évaluation comme problématique du sens) qui ambitionnent d'articuler explicitement des orientations épistémologiques

et des débats paradigmatiques, des modèles disciplinaires et interdisciplinaires et des modèles dits « locaux », c'est-à-dire spécifiques à un domaine ou une équipe de recherche. Une grande complexité se dégage de cette conceptualisation et, au-delà de ce cas particulier, s'impose alors une impression d'éclatement des propositions (Figari, 2006) ou ce que Dortier (2006) nomme une « intelligence dispersée » dans les sciences sociales de ces dernières années : « Notre pensée est éclatée, dispersée, en une myriade de connaissances locales qu'aucune grille de lecture globale ne permet de rassembler en un tout cohérent » (p. 22).

Les méthodes d'évaluation ou modèles d'action

Le dernier niveau de conceptualisation que nous distinguons relève des méthodes d'évaluation ou modèles d'action qui se caractérisent par l'énonciation de principes prescriptifs, parfois quasi a-théoriques. Nous pensons que c'est le cas du « modèle » CIPP (*Context, Input, Processus, Product*) très connu de Stufflebeam *et al.* (1980), ou de celui que De Ketele (1993) appelle le « paradigme de l'évaluation au service de la décision » dont la finalité se veut essentiellement pragmatique : les décisions les plus adéquates doivent pouvoir être prises au moyen de la démarche évaluative. Plus généralement, il est souvent reproché à l'évaluation de se réduire à des prescriptions normatives ou à des applications mécaniques d'instruments et de « façons de faire » (conception « applicationniste » de l'évaluation, Morrissette, Loye & Legendre, 2012) sans fondements théoriques explicites ni mises en doute critiques. Notre point de vue n'est évidemment pas de nier les dimensions pratiques et idéologiques de l'évaluation, mais de les penser en interaction avec leurs orientations dites paradigmatiques et les théorisations, y compris méthodologiques, qui les fondent.

Enjeux d'une réflexion critique sur la modélisation pour l'évaluation

Ainsi, il ressort de la littérature une grande diversité de conceptualisations de l'évaluation en éducation, avec des usages terminologiques différents, voire opposés. Pour Figari (2006), cette disparité reflète l'état d'esprit des travaux des années 1980 qui font coexister des concepts d'origines diverses et qui contribuent à décrire de manière de plus en plus complexe les processus d'évaluation. On s'étonnera pourtant que, dans la période récente, les théorisations de l'évaluation n'aient pas donné lieu à davantage de travaux visant à construire collectivement une plus grande cohérence du champ. Afin de contribuer modestement à cette visée, un des enjeux de cet ouvrage est déjà de donner à voir et à débattre, dans une perspective épistémologique, quelques *modélisations* élaborées par des chercheurs

d'orientations différentes lorsque leur objet porte sur l'évaluation dans un contexte éducatif ou de formation.

Par rapport au modèle, préconstruit par nature, la modélisation, qu'elle soit mathématique ou non, constitue une procédure active, appliquée à la préparation d'une évaluation précise et locale, en relation avec une problématique qui, elle, est générale. Selon Figari (1998), la modélisation consiste en une représentation abstraite des relations (de cause à effet, de comparaison, de répétition de phénomènes) existant entre les éléments soumis à évaluation et les facteurs qui les influencent. La modélisation revient à construire un cadre d'interprétation cohérent et fondé sans lequel l'évaluation devient subjective et arbitraire (Machado, 2007). Pour Le Moigne (1977/2006), « la modélisation postule *a priori* non seulement la pluralité des modèles concevables d'un même phénomène, mais surtout la pluralité des *méthodes de modélisation* [souligné par l'auteur] » (p. 14). De fait, on constate dans la littérature spécialisée en évaluation qu'il est plus souvent question de « modèle », plus rarement de « modélisation » et encore moins de conceptualisation des différentes formes de modélisation opérées dans une perspective explicite de mise à distance épistémologique de l'évaluation. C'est donc là l'un des enjeux de cet ouvrage par rapport à l'évaluation, conçue à la fois comme une activité de recherche et une pratique sociale située.

ORGANISATION DE L'OUVRAGE

L'ouvrage réunit un ensemble de travaux scientifiques autour de trois axes, correspondant chacun à une partie, qui visent à organiser les conditions du débat sur les questions épistémologiques soulevées dans cette introduction : 1) Les activités d'évaluation : quels savoirs produits ? 2) Les activités d'évaluation : quelles modélisations mises en œuvre ? 3) D'une épistémologie absente à une épistémologie en débat.

Les activités d'évaluation : quels savoirs produits ?

Cette première partie de l'ouvrage propose une réflexion critique sur les savoirs élaborés par le moyen de l'activité évaluative et/ou de la recherche sur cette activité. Schématiquement, deux catégories de savoirs peuvent être distinguées : 1) le savoir produit par la pratique évaluative (d'un apprenant, d'un enseignant/formateur, d'un consultant, d'un chercheur, etc.) qu'il est possible de qualifier de local, contextualisé, circonstancié, mouvant, peu généralisable ; 2) le savoir produit par la recherche sur des phénomènes relevant d'une problématique évaluative, susceptible d'être plus représentatif

et possiblement transférable et généralisable. Par rapport aux publications spécialisées sur les types de savoirs qui caractérisent les sciences de l'éducation (voir Hofstetter & Schneuwly, 2009), la spécificité de notre démarche est de questionner la relation entre ces différents niveaux ou types de savoirs et les pratiques qui relèvent d'une problématique explicitement évaluative. Quel est le statut de l'évaluation (en tant que praxis et/ou praxéologie, si on reprend les termes de Bedin [2008] cités *supra*) par rapport aux savoirs qu'elle produit ? Quelle rupture ou possible circularité les auteurs conçoivent-ils entre ces différents types de savoirs ? En prenant appui sur des corpus de données et des résultats de recherches, trois chapitres tentent de répondre à ces questions et, ce faisant, mettent en débat les savoirs produits par différentes démarches évaluatives.

La contribution de *Joëlle Morrisette, Lucie Mottier Lopez et Walther Tessaro* compare deux recherches collaboratives sur les pratiques d'évaluation formative d'enseignants du primaire. Elle examine la nature des savoirs produits au regard notamment du rôle joué par le cadre théorique dans les processus d'analyse et d'interprétation collective des données recueillies. La première recherche défend une posture foncièrement inductive, alors que la deuxième s'inscrit dans une approche inductive-déductive. Une interfécondation entre les savoirs pratiques et les savoirs scientifiques est postulée, amenant les auteurs à insister sur la portée des « savoirs [collectifs] négociés [...] susceptibles [...] d'offrir des contributions significatives au domaine de l'évaluation des apprentissages des élèves en classe » (p. 42).

Le chapitre de *Dominique Lafontaine et Christian Monseur* s'intéresse aux savoirs produits par les grandes enquêtes internationales, tel PISA⁵. Le projet des auteurs est de montrer en quoi les choix théoriques et méthodologiques contraignent les résultats de l'évaluation des acquis des élèves, et d'interroger la forme des savoirs produits et leurs usages sociaux par les acteurs de la scène éducative. « La demande est souvent forte pour pousser le chercheur à dépasser les constats et à trancher des questions de politique éducative sur la base de son expertise des enquêtes internationales » (pp. 58-59). Les auteurs soulignent le rôle délicat de « garde-fou » du chercheur en évaluation face à des questions qui relèvent de l'éthique et de l'axiologie.

La contribution de *Claire Tourmen et Patrick Mayen* oriente le regard sur l'épistémologie de l'évaluation à travers celle(s) des évaluateurs. Partant de l'analyse du travail des évaluateurs dans des jurys de validation des acquis de l'expérience et dans des épreuves en situation professionnelle dans l'enseignement secondaire agricole, les auteurs interrogent le rapport des évaluateurs aux phénomènes évalués, mais également la « théorie de la construction du savoir » qu'ils convoquent et ce qu'ils en font dans l'action et pour l'action. Les auteurs montrent que l'activité évaluative est « structurée par l'expérience des

5. Programme international pour le suivi des acquis des élèves.

phénomènes évalués [...] qui nourri[t] le processus d'évaluation et perme[t] de s'en distancier dans un même mouvement » (pp. 74-75).

À travers ces trois premières contributions, différents savoirs que l'activité d'évaluation est susceptible de produire au regard des modèles théoriques de référence et des choix méthodologiques effectués se dessinent. On s'éloigne déjà ici de la doxa selon laquelle la fonction de l'évaluation ne consisterait qu'à produire des « jugements de valeur » (normatifs), pour se diriger vers le postulat d'une « théorie pragmatique » de l'évaluation qui, dans sa conceptualisation, prend pleinement en compte les contextes expérimentiels avec lesquels l'évaluation prend sens.

Les activités d'évaluation : quelles modélisations mises en œuvre ?

La deuxième partie de l'ouvrage a pour projet d'exposer différentes modélisations opérées dans des recherches en évaluation et d'engager une conceptualisation de celles-ci, en explicitant par exemple les liens entre la formulation conceptuelle et la situation empirique concernée, en soulignant les possibles reconfigurations en fonction de la spécificité du phénomène évaluatif étudié, en nommant les phases constitutives de la modélisation, en l'illustrant par une schématisation. Le projet est ambitieux : comment les auteurs se situent-ils par rapport aux processus dynamiques de modélisation qu'ils ont mis en œuvre dans les recherches qu'ils présentent ? Quel regard épistémologique portent-ils sur leurs choix ? Les trois premières contributions de cette seconde partie s'inscrivent dans le contexte scolaire, les deux suivantes en formation continue d'enseignants.

Le chapitre de *Pierre Merle* interroge la modélisation des pratiques d'évaluation et de notation des compétences des élèves à partir de trois perspectives : d'abord l'approche docimologique, puis les modèles sociologiques déterministe et interactionniste. L'auteur opte en faveur du modèle interactionniste de l'évaluation, présentant selon lui une pertinence supérieure au modèle déterministe, mais sans pour autant rendre ce dernier caduc. L'auteur effectue, à partir de données empiriques, la démonstration du « pouvoir interprétatif » du modèle qu'il privilégie et de son épistémologie, défendant le point de vue qu'« il n'existe pas de rupture définitive et infranchissable [...] entre la connaissance et l'expérience des acteurs » (p. 94).

Le chapitre de *Pierre Valois, Claude Houssemand et Astrid de Leeuw* présente les étapes du processus de modélisation d'un instrument de mesure, selon une notion élargie du concept de mesure et de son épistémologie dans les sciences humaines et sociales. Les auteurs s'appuient sur l'exemple de la construction et de la validation d'un questionnaire d'attitudes en éducation servant à mesurer les attitudes des enfants du primaire envers une

alimentation équilibrée. Ils discutent la relation avec les modèles explicatifs et prédictifs qui donnent sens à la mesure, « en adoptant la position selon laquelle la mesure consiste en un exercice méthodologique » (p. 111), ainsi que la combinaison nécessaire entre des études qualitatives et quantitatives.

La contribution de *Dany Laveault* porte sur l'évaluation des apprentissages des élèves dans une perspective formative. L'auteur a pour objectif de présenter « comment les modèles de [la] régulation [...] sont associés aux enjeux de l'évaluation formative chez les chercheurs francophones et à ceux de l'évaluation-soutien d'apprentissage chez les chercheurs anglo-saxons » (p. 115). L'exploitation de modèles, en particulier de l'autorégulation, est considérée comme indispensable pour expliquer comment l'évaluation et l'autoévaluation sont susceptibles de soutenir la progression des apprentissages des élèves. L'auteur plaide pour une ouverture à des stratégies de régulation émotionnelle, montrant, ce faisant, la transformation de sa modélisation au fil des avancées théoriques et résultats de recherches.

Le chapitre de *Michel Vial* conceptualise les processus de modélisation de l'évaluation en prenant pour exemple des pratiques d'accompagnement d'enseignants. Le projet de l'auteur est de mettre au jour les liens entre les « modes de pensée » explicitement ancrés dans une réflexion épistémologique et les différents modèles de l'évaluation. La modélisation pour cet auteur s'élabore dans un mouvement dynamique et émergent, en tant que « processus heuristique fait de méandres, de retours en arrière » (p. 142). Chaque mode de pensée est considéré comme « un élément organisateur de la pratique, un analyseur qui permet de donner sa place à une épistémologie profane » (p. 143) ; plusieurs modélisations existent alors nécessairement en fonction des histoires singulières de chacun.

Enfin, le chapitre de *Pedro Rodrigues et Eusébio André Machado* expose deux « trajets de modélisation » de la participation des acteurs aux processus d'évaluation. À partir de deux exemples de recherches dans le contexte de la formation d'enseignants, les auteurs conceptualisent les processus de modélisation avec l'intention de montrer qu'ils « font partie du processus générique de recherche qui passe par la confrontation (dialectique) d'un modèle théorique, hypothétique, avec les données provenant d'une situation réelle » (p. 159). La démonstration des auteurs veut se situer sur un plan métathéorique, afin de tenter d'objectiver la méthode de construction d'une modélisation d'une réalité donnée portant sur une problématique évaluative.

D'une façon générale, cette deuxième partie de l'ouvrage défend le postulat que la validité et le sens des résultats produits par la recherche en évaluation sont intimement liés aux modélisations effectuées, quel que soit d'ailleurs l'utilisation (à des fins scientifiques, pratiques, administratives, politiques...) de ces résultats. Les contributions tendent à montrer la nécessité de penser en termes de système les différents niveaux de conceptualisation

dégagés *supra*, entre ancrages paradigmatiques, modèles (scientifiques ou profanes) et expériences situées.

D'une épistémologie absente à une épistémologie en débat

Les auteurs des deux premières parties de l'ouvrage ont engagé, de façon plus ou moins explicite et chacun à leur manière, une réflexion épistémologique en questionnant les savoirs produits par la démarche évaluative, les modèles de référence et modélisations opérées dans le cadre de leurs recherches. Dans la littérature, on a observé que, finalement, peu de dissertations épistémologiques existaient sur la question. Cette dernière partie de l'ouvrage vise donc à examiner plus directement l'idée d'une « épistémologie de l'évaluation ». Au regard de l'état actuel de développement des travaux dans le champ de l'évaluation en éducation et des nombreux bilans qui témoignent de la vivacité de ce champ scientifique, peut-on aujourd'hui envisager, avec toute la prudence qui s'impose, l'hypothèse d'une épistémologie spécifique ? Cette partie de l'ouvrage présente trois contributions qui ont pour particularité de défendre des points de vue contrastés sur la question, puis un chapitre conclusif qui ose évoquer une épistémologie spécifique à l'évaluation, tout en examinant et discutant les apports des différents chapitres.

La contribution de *Lise Demailly* ouvre le débat en exposant un positionnement critique sur les enjeux épistémologiques de la méta-évaluation, définie comme un travail de réflexivité sur la qualité des évaluations, quelles qu'elles soient, dans le but de les modifier, voire de les améliorer. « L'évaluation n'est pas une pratique sociale seulement stratégique, elle est guidée par des valeurs de vérité, de connaissance, de production de savoirs sur le réel » (p. 177). L'auteure affirme cependant que la question épistémologique ne peut pas être traitée de façon autonome, sans prendre aussi en considération les enjeux politiques et institutionnels constitutifs de toute (méta-)évaluation.

La contribution de *Linda Allal* s'insurge contre l'idée de concilier dans un même paradigme de référence l'ensemble des pratiques d'évaluation en éducation. L'auteure effectue sa démonstration en distinguant deux grands champs : celui de l'évaluation des caractéristiques des personnes et celui de l'évaluation des dispositifs de formation. La spécificité de chacun d'eux, au regard des rôles des acteurs, des objets évalués et des finalités visées par l'évaluation, imposerait une nécessaire différenciation de leur ancrage paradigmatique. « Les deux champs doivent cohabiter, voire s'articuler dans le fonctionnement des systèmes éducatifs, même si chacun possède sa dynamique propre non réductible à celle de l'autre champ » (p. 182).

À l'inverse, le chapitre de *Jean-Marie De Ketele* tente le pari de la conceptualisation d'un paradigme unificateur, à partir de la détermination

de quatre modèles (ou paradigmes princeps) et de l'exposition de leurs fondements conceptuels et idéologiques : l'évaluation intuitive pragmatique, l'évaluation mesure, l'évaluation de la maîtrise des apprentissages et l'évaluation pour les apprentissages ou « soutien d'apprentissage », selon les termes de Laveault. L'objectif de l'auteur est de « voir si un paradigme unificateur, à la fois d'élargissement (voir les travaux de Cardinet) et d'intégration, permet de mieux comprendre le développement des travaux sur l'évaluation » (p. 196).

La contribution de *Gérard Figari* clôt l'ouvrage en discutant les réponses d'ordre épistémologique qui ont été formulées par les contributeurs autour des questions posées. « C'est donc, tout d'abord, le lieu de faire l'inventaire des positions, des oppositions et des propositions avant de mesurer le cheminement d'une théorisation : celle consistant à identifier et à schématiser les dimensions épistémologiques de l'évaluation en éducation, telle qu'elle est *pensée* et non pas seulement *pratiquée* aujourd'hui » (p. 212). L'enjeu de ce chapitre consiste alors à dégager des configurations possibles d'une épistémologie de l'évaluation, en tentant d'adopter une vision en surplomb dans un débat qui est destiné, comme le conclut l'auteur, à être poursuivi bien au-delà de cet ouvrage.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. (2009). Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire. In L. Mottier Lopez & M. Crahay (Éd.), *Évaluations en tension : entre la régulation des apprentissages et le pilotage des systèmes* (pp. 29-45). Bruxelles : De Boeck.
- Allal, L., Cardinet, J. & Perrenoud, P. (Éd.). (1979). *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.
- Amigues, R. & Zerbato-Poudou, M.-T. (1996). *Les pratiques scolaires d'apprentissage et d'évaluation*. Paris : Dunod.
- Bedin, V. (2008, juin). *Pour une praxis de l'évaluation*. Conférence donnée dans le cadre du séminaire Lectures épistémologiques autour de l'évaluation, Toulouse. Consulté le 3 mars 2012 dans <http://w3.crefi.univ-tlse2.fr/sommaire/seminaire/lectureepisteval2.pdf>
- Blais, J.-G. (2005). L'obligation de résultats à la lumière de l'interaction entre le quantitatif et le social. In C. Lessard & P. Meirieu (Éd.), *L'obligation de résultats en éducation. Évolution, perspectives et enjeux internationaux* (pp. 123-144). Bruxelles : De Boeck.
- Blais, J.-G., Larose, F., Laurier, M., Lessard, C., Rousseau, C., Dupuis, P. & Proulx, J.-P. (1999). Le palmarès des écoles en question. *Mesure et évaluation en éducation*, 22(1), 1-20.
- Bonboir, A. (1972). *La docimologie*. Paris : Presses Universitaires de France.

- Bonniol, J.-J. & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles : De Boeck.
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris : Presses Universitaires de France.
- Cardinet, J. (1990). Évaluation interne, externe ou négociée ? In *Hommage à Jean Cardinet* (pp. 139-156). Fribourg : Delval.
- Cowley, P. & Marceau, R. (2000). *Bulletin des écoles secondaires du Québec*. Montréal : Institut économique de Montréal et The Fraser Institute.
- De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, 103, 59-80.
- De Ketele, J.-M. (2001). Évolution des problématiques issues de l'évaluation formative. In G. Figari & M. Achouche (Éd.), *L'activité évaluative réinterrogée : regards scolaires et socioprofessionnels* (pp. 102-108). Bruxelles : De Boeck.
- De Landsheere, G. (1971). *Évaluation continue et examens. Précis de docimologie*. Bruxelles : Labor.
- Demailly, L. (2006). En Europe : l'évaluation contre la crise des systèmes scolaires, l'évaluation en crise. *Éducation et formations*, 17, 105-120.
- Dortier, J.-F. (2006). L'intelligence dispersée. *Sciences humaines*, 167, 22-23.
- Figari, G. (1998). Évaluation, science des conjectures : regard critique sur le texte de Le Poulitier. In J. Py, A. Somat & J. Baillé (Éd.), *Psychologie sociale et formation professionnelle*. Rennes : Presses Universitaires de Rennes.
- Figari, G. (2006, septembre). *Pour une épistémologie de l'évaluation*. Conférence donnée au 19^e colloque de l'ADMEE-Europe, Luxembourg.
- Fournet, M. (2008, juin). *Imaginaires et projets dans l'évaluation*. Conférence donnée dans le cadre du séminaire Lectures épistémologiques autour de l'évaluation, Toulouse. Consulté le 3 mars 2012 dans <http://w3.crefi.univ-tlse2.fr/sommaire/seminaire/lectureepisteval2.pdf>
- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA : Sage Publications Inc.
- Hofstetter, R. & Schneuwly, B. (Éd.). (2009). *Savoirs en (trans)formation : au cœur des professions de l'enseignement et de la formation*. Bruxelles : De Boeck.
- Kuhn, T. S. (1970). *La structure des révolutions scientifiques*. Paris : Champs Flammarion.
- Le Moigne, J.-L. (1977/2006). *La théorie du système général, théorie de la modélisation*. Paris : Presses Universitaires de France.
- Le Moigne, J.-L. (1999). *La modélisation des systèmes complexes*. Paris : Dunod.
- Machado, E. A. (2007). *Évaluation et participation. Une étude sur le rôle des acteurs dans l'évaluation de la formation continue*. Thèse de doctorat en Sciences de l'éducation, Universidade do Minho.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.

- Morrisette, J., Loye, N. & Legendre, M.-F. (2012). Autour du savoir enseigner des enseignantes et des enseignants en salle de classe. *Mesure et évaluation en éducation*, 34(3), 1-4.
- Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 3, 465-482.
- Mouloud, N. (2002). Modèle. In *Encyclopaedia universalis* (Vol. 15, pp. 294-295). Paris : Encyclopaedia universalis.
- Mucchielli, A. (1996). *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*. Paris : Armand Colin.
- Nagels, M. & Vourch, C. (2009). De l'évaluation, considérée comme une relation de service. *Questions vives, évaluer les enseignants et les formateurs, comment, pourquoi, pour quoi ?* 6(12), 5-59.
- Pelletier, G. (2009). L'évaluation-conseil en milieu universitaire : une analyse de pratiques internationales. In V. Bedin (Éd.), *L'évaluation à l'université, évaluer ou conseiller ?* (pp. 99-114). Rennes : Presses Universitaires de Rennes.
- Piéron, H. (1963). *Examens et docimologie*. Paris : Presses Universitaires de France.
- Rege Colet, N. (2009). La place de l'évaluation formative dans le développement professionnel des enseignants-chercheurs. In V. Bedin (Éd.), *L'évaluation à l'université, évaluer ou conseiller ?* (pp. 275-294). Rennes : Presses Universitaires de Rennes.
- Rodrigues, P. (2006). Logiques d'évaluation en relation avec des conceptions élargies du monde et de la connaissance. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 193-201). Paris : L'Harmattan.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O. & Provus, M. M. (1980). *L'évaluation et la prise de décision en éducation*. Victoriaville : N.H.P.

LES ACTIVITÉS D'ÉVALUATION : QUELS SAVOIRS PRODUITS ?

La production de savoirs négociés dans deux recherches collaboratives sur l'évaluation formative

**Joëlle Morrissette*,
Lucie Mottier Lopez** & Walther Tessaro****

***Université de Montréal**

****Université de Genève**

Différents modèles d'intelligibilité de l'évaluation des apprentissages sont aujourd'hui reconnus, proposant des types de savoirs dont le sens et la légitimité sont liés à l'architecture des investigations conduites, et donc aux systèmes de représentations qu'elles convoquent ainsi qu'aux traditions de recherche dans lesquelles elles s'inscrivent. Parmi les recherches portant plus précisément sur l'évaluation des apprentissages des élèves, un nombre croissant se réclament d'une approche collaborative selon laquelle la participation des acteurs de terrain est requise dans l'élaboration de savoirs produits par et sur l'évaluation. Relevant d'une épistémologie de l'agir, cette approche repose sur l'hypothèse selon laquelle une interfécondation est possible entre les communautés de pratique professionnelle et les communautés scientifiques (Desgagné, 2001).

Dans ce chapitre, nous présenterons deux recherches collaboratives portant sur des pratiques d'évaluation formative d'enseignant(e)s du primaire. L'une d'elles étudie le savoir-faire d'enseignantes d'une école située dans la région de Québec (Morrissette, 2010a). S'inscrivant dans l'héritage de la « tradition de Chicago » (Chapoulie, 2001), elle s'intéresse à ce qui constitue, du point de vue des praticiennes, des démarches d'évaluation formative. L'autre, réalisée dans le canton de Genève en Suisse, porte sur les régulations qui relèvent des pratiques d'évaluation formative d'enseignants dans des résolutions de problèmes mathématiques (Mottier Lopez *et al.*, 2010).

En s'appuyant sur une perspective située, elle s'intéresse aux dynamiques interactives de la régulation formative et aux significations coélaborées à leur propos entre les enseignants et la chercheure.

Ces investigations ont plusieurs points communs, dont le fait de partager un même objet d'étude : l'évaluation formative en salle de classe. Dans le monde francophone, l'évaluation formative est étroitement associée au concept de « régulation » des apprentissages et des processus d'enseignement, se démarquant ainsi de la notion de *feedback* qui est au cœur des travaux anglophones. De nombreuses conceptualisations ont été proposées en fonction des différents ancrages théoriques et épistémologiques des auteurs (voir à ce propos les synthèses d'Allal & Mottier Lopez [2005] et de Morrissette [2010b]). Dans le monde anglophone, c'est l'expression *formative assessment* qui a aujourd'hui tendance à s'imposer lorsqu'il est question des apprentissages des élèves à l'école ou encore *assessment for learning* (Black & Wiliam, 2006)¹, voire *assessment as learning* (Dann, 2002) qui renforce l'idée que l'évaluation peut être source d'apprentissages. Ces différentes conceptions se démarquent d'un modèle de l'évaluation qui serait presque exclusivement décliné en démarches formelles, standardisées et forcément instrumentées.

Les investigations dont il est question dans cette contribution s'intéressent plutôt aux dimensions interactives de la pratique évaluative, dimensions encore peu explorées dans les théorisations de l'évaluation formative. Dans la mouvance des perspectives socioculturelles, l'évaluation formative est envisagée comme une pratique sociale, discursive et située, se réalisant au travers de processus de coconstruction de sens imbriqués aux activités quotidiennes et médiatisés par la culture et le langage (Morrissette, 2009 ; Mottier Lopez, 2006). Ces investigations rompent avec le rapport applicationniste que nombre de chercheurs entretiennent à la pratique (évaluative) des enseignants, en leur reconnaissant une compétence d'acteurs. En s'appuyant sur les théories de Schön (1983/1994) et de Giddens (1987), elles postulent que tous les praticiens développent des savoirs pratiques au fil de l'expérience quotidienne, qu'ils jouissent d'une certaine marge de manœuvre et disposent de ressources pour agir. En outre, leur réflexivité, c'est-à-dire leur capacité à interpréter leur agir, conduit à déployer des dispositifs collaboratifs visant à intégrer leur compréhension en contexte dans les modélisations de la pratique.

Si les deux recherches collaboratives adhèrent au postulat selon lequel il n'y a pas d'évaluation formative sans *modèle*, qu'il soit explicite ou implicite, elles diffèrent toutefois sur différents points, notamment sur le rôle joué par le cadre théorique dans les processus d'analyse et d'interprétation collective des informations recueillies. La place de la théorie et des outils d'analyse, ainsi que leur impact sur la modélisation des pratiques étudiées,

1. Voir les contributions de Laveault et De Ketele dans le présent ouvrage.

ne sont pas simples dans les recherches qui se veulent vraiment collaboratives. En effet, comme le signalent Bourassa, Bélair et Chevalier (2007), le principe épistémologique de ce type de recherche exigerait que l'analyse des informations recueillies soit effectuée avec les acteurs de terrain dans l'optique d'une théorisation coconstruite de la pratique. Les deux recherches avaient pour enjeu de coanalyser et de co-interpréter l'expérience des praticiens en lien avec la mise en œuvre de l'évaluation formative. La première recherche a procédé de manière *inductive*, c'est-à-dire sans modèle de l'évaluation formative explicite proposé aux enseignants, en raison de son ancrage dans la « tradition de Chicago » qui incite à saisir et restituer, de manière compréhensive, les significations des activités et des expériences des acteurs sociaux en relation avec leurs contextes de réalisation. Quant à la seconde recherche, elle a fait appel à une démarche *inductive-déductive* qui respecte la cohérence interne des données tout en utilisant des descripteurs issus de cadres théoriques à disposition dans la littérature de recherche (Balslev & Saada-Robert, 2002).

À partir de cette différence théorique et épistémologique, l'objectif de ce chapitre est d'examiner comment les modèles participent de la production de savoirs sur l'évaluation formative dans le cadre de ces deux recherches et de discuter la nature des savoirs produits en tant que contributions au domaine de l'évaluation formative.

UNE RECHERCHE IDENTIFIANT LE SAVOIR DE TERRAIN EN RELATION AVEC LA MISE EN ŒUVRE DE L'ÉVALUATION FORMATIVE

La première recherche a théorisé le savoir de terrain d'un groupe de cinq enseignantes volontaires du primaire² à partir de leurs « manières de faire »³ l'évaluation formative des apprentissages. Pendant une période s'étalant sur cinq mois, elles ont été invitées à participer à une démarche de coconstruction de savoirs professionnels composée de trois types d'activités réalisés en alternance. De fait, elles ont examiné trois bandes vidéo produites dans leur classe afin d'identifier d'éventuels épisodes d'évaluation formative ; elles ont ensuite explicité ceux-ci lors de trois entretiens individuels ; enfin, elles ont rapporté ces épisodes à leurs pairs lors de cinq entretiens de groupe. Cette recherche, d'une part, a formalisé les modèles implicites sous-jacents à leurs démarches d'évaluation formative à partir de l'explicitation de leurs pratiques respectives, suivant la compréhension contextuelle qu'elles en

2. Il s'agit d'une enseignante de 6^e année, une de 5^e année et trois de 4^e année. Elles ont entre 2 et 25 ans d'expérience professionnelle.

3. « Manières de faire » (de Certeau, 1990) ne réfère pas au sens commun ; il s'agit d'une expression chargée théoriquement, renvoyant notamment au pragmatisme de Dewey dans lequel s'ancre la « tradition de Chicago ».

ont. D'autre part, elle a conduit à une théorisation a posteriori du territoire symbolique de l'évaluation formative au regard de différentes zones du savoir enseignant (Morrissette, 2011).

Un modèle qui organise la pratique : l'exemple d'une routine d'évaluation formative

Lors des entretiens individuels, les enseignantes ont présenté les épisodes repérés et ont explicité les démarches mises en œuvre dans le cadre de l'interaction de la classe. Le protocole de rétroaction vidéo relève d'une « réflexion partagée » (Tochon, 1996), soit une perspective fondée sur une théorie de coconstruction sociale et culturelle des connaissances. Dans ce cadre, l'enregistrement vidéo facilite le travail de reconstruction de l'expérience. C'est à partir de ces entretiens qu'ont été inférés les modèles qui structurent les démarches d'évaluation formative des enseignants. À titre illustratif, nous présentons celui d'une enseignante de 6^e année (élèves de 10-11 ans) qui renvoie à une routine d'évaluation formative organisant sa pratique hebdomadaire.

C'est à la suite du deuxième entretien individuel que l'enseignante a sélectionné quelques extraits vidéo qui lui permettent de scénariser sa pratique, en recomposant ce qui constitue, de son point de vue, chacune des quatre étapes de sa routine hebdomadaire d'évaluation formative.

Tableau 1 : Une routine d'évaluation formative

<i>Étape 1</i>	Enseignement de nouveaux savoirs en groupe
<i>Étape 2</i>	Rétroaction individuelle lors des devoirs réalisés en classe
<i>Étape 3</i>	Rétroaction collective autour des principales difficultés observées
<i>Étape 4</i>	Rétroaction individuelle lors de la « mise à jour » des devoirs par les élèves

Les étapes qui structurent la routine d'évaluation formative se déroulent entre le lundi et le jeudi, les dernières périodes horaires de la journée étant consacrées à la réalisation et à la correction des devoirs en classe expliqués dès la première période du matin.

L'étape 1 correspond au moment où l'enseignante présente de nouveaux savoirs à l'ensemble du groupe-classe, lors de la première période du matin. Cette étape est elle-même précédée d'une évaluation formative l'ayant amenée à identifier les savoirs que les élèves doivent s'approprier pour développer une compétence donnée. Elle correspond aux phases de préparation et de réalisation d'une démarche didactique classique. Validant au fur et à mesure de son enseignement la compréhension des élèves en

leur posant fréquemment des questions, l'enseignante dit anticiper déjà certaines difficultés qui pourraient se poser pendant cette première étape.

L'étape 2 concerne la dernière période horaire de la journée. Les élèves réalisent des devoirs en classe qui nécessitent la mobilisation des nouveaux savoirs enseignés. Ils travaillent seuls ou en équipe, selon les cas, et interpellent l'enseignante au besoin : certains sollicitent des explications supplémentaires, d'autres des validations du travail effectué. Les échanges avec chacun des élèves, notamment le type de questions posées et les traces des démarches entreprises, lui permettent, selon elle, de dépister les raisons des erreurs et de comprendre ce qui fait obstacle. Cette étape 2 vise à déterminer avec chacun des élèves une stratégie différenciée afin de les aider à surmonter la difficulté d'apprentissage.

Le lendemain matin, lors de *l'étape 3*, l'enseignante procède à une rétroaction collective en fonction des principales difficultés qu'elle a observées la veille. À cette étape, elle mise sur un processus de questionnement continu pour faire contribuer le groupe à l'éclaircissement des notions enseignées, à l'identification des démarches (in)fructueuses, etc. Cette étape se réalise sur un mode interactif, l'intention de l'enseignante étant que l'évaluation formative se coconstitue au sein d'échanges entre élèves qu'elle encadre et soutient.

L'étape 4 se réalise à la fin de cette deuxième journée. C'est l'étape appelée « mise à jour » par l'enseignante : chaque élève retravaille les tâches amorcées la veille en fonction de l'éclairage apporté par le travail collectif effectué le matin même, et commence une autre série de devoirs. Cette période horaire se déroule sur un mode très similaire à celui de l'étape 2 : les obstacles d'apprentissage et les stratégies visant à les surmonter sont codéfinis par l'enseignante et l'élève.

Ainsi, pour cette enseignante, l'évaluation formative est modélisée par une routine d'enseignement/apprentissage qui combine des moments de coconstruction de sens collectivement et de façon plus individualisée autour des savoirs à s'approprier. La rétroaction négociée et l'intervention différenciée jouent un rôle prépondérant dans la « manière de faire » de l'enseignante. Ce modèle d'une pratique d'évaluation formative routinière, ancrée dans une façon particulière d'aménager la grille horaire, semble structurer son action en classe, son savoir-faire du quotidien qui implique alors une gestion de classe déterminée.

Une théorisation du savoir enseignant en matière d'évaluation formative

En alternance avec les entretiens individuels, des entretiens de groupe ont été tenus, s'appuyant sur le postulat selon lequel l'interprétation que les enseignantes font de leur expérience s'inscrit dans un ensemble de relations avec

d'autres acteurs, impliqués dans une expérience commune de soutien aux apprentissages. Cette perspective invite à considérer que la mise en œuvre de l'évaluation formative relève, en partie du moins, d'un savoir-faire collectif. Ainsi, c'est lors des entretiens de groupe que les épisodes coanalysés par chacune des enseignantes et la chercheure ont été rapportés et commentés, permettant à la réflexivité de s'approfondir dans l'intersubjectivité.

UNE REPRÉSENTATION SCHÉMATIQUE COMME LEVIER À LA COCONSTRUCTION DE L'OBJET

Au cours des premiers entretiens de groupe, la chercheure est restée très en retrait, se limitant à inviter les enseignantes à rapporter un épisode identifié sur les bandes vidéo et à distribuer les tours de parole si nécessaire. Les praticiennes avaient donc une large marge de manœuvre pour cheminer vers l'explicitation de leurs pratiques, en fonction de ce qu'elles retenaient comme étant pertinent collectivement relativement à la pratique de l'évaluation formative. Par ailleurs, sollicitée à plusieurs reprises par les enseignantes désireuses d'obtenir une forme de rétroaction sur leurs pratiques, la chercheure a introduit une représentation schématique des contributions respectives lors du troisième entretien (Morrisette, 2010a). Élaborée à l'aide de catégories dégagées à partir de la question : « De quoi parlent-elles ici ? », cette représentation semble avoir répondu au besoin exprimé par les enseignantes, tout en permettant de tester la validité des interprétations en continu (van Campenhoudt, Chaumont & Franssen, 2005). À chaque entretien subséquent, une version augmentée de la représentation schématique était proposée aux enseignantes qui pouvaient alors (in)valider ou préciser les interprétations de la chercheure. Reconduisant leurs propres termes, cette représentation est devenue pour les praticiennes un artefact aménageable sur lequel prendre appui pour négocier le sens de leurs « manières de faire », servant de levier à la coconstruction de l'objet. Ainsi, c'est le sens coélaboré par les participantes en continu pendant la démarche de terrain et traduit dans cette représentation qui a constitué une grille théorique *en émergence* pour instrumenter les échanges intersubjectifs.

UNE ANALYSE PERMETTANT DE DISTINGUER LES « MANIÈRES DE FAIRE » PARTAGÉES, ADMISES ET CONTESTÉES

Les débats d'idées lors des entretiens de groupe ont incité à déployer une analyse interactionniste⁴ sur les cinq verbatim, ciblant les « négociations conversationnelles » lorsque les « manières de faire » des unes et des autres étaient rapportées et discutées. C'est ainsi qu'ont été distinguées :

4. L'approche analytique développée dans le cadre de cette recherche se fonde sur l'hypothèse selon laquelle le sens construit lors des entretiens est indissociable de la façon dont il est produit dans l'interaction entre les collaboratrices. Pour en connaître les détails, les lecteurs sont invités à consulter Morrisette (2011).

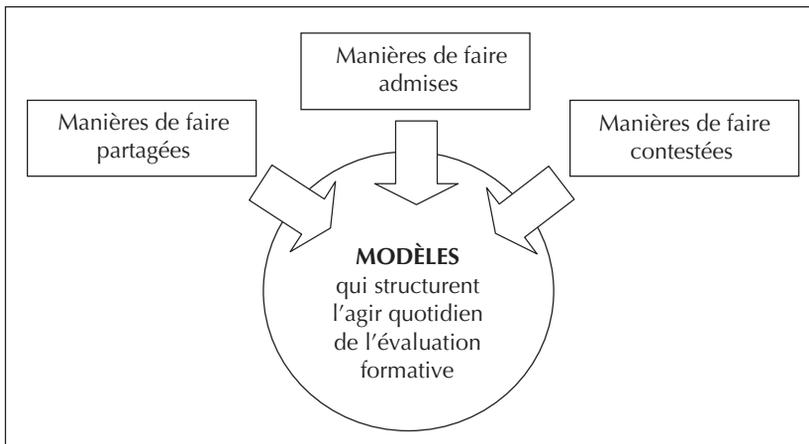
- Des « manières de faire partagées », soit celles qui ont suscité des processus de reconnaissance mutuelle dans l'interaction, des constructions consensuelles ;
- Des « manières de faire admises », soit celles qui ont fait l'objet d'une explicitation par une enseignante, sans marque d'adhésion ou d'objection de la part de ses pairs ;
- Des « manières de faire contestées », soit celles qui ont suscité des désaccords persistants entre les enseignantes, n'étant pas reconnues d'emblée comme des pratiques d'évaluation formative.

À titre d'exemple, la routine d'évaluation formative explicitée précédemment a été identifiée comme une « manière de faire admise » : étant propre à l'enseignante, elle a été reconnue par ses pairs comme une pratique légitime – voire inspirante – mais néanmoins non partagée.

UNE THÉORISATION DU TERRITOIRE SYMBOLIQUE DE L'ÉVALUATION FORMATIVE AUTOUR DE TROIS ZONES DE SAVOIR

La posture inductive endossée par la chercheuse invite à mobiliser de nouveaux cadres théoriques pour donner un sens à la conceptualisation émergente. Ainsi, en lien avec l'objectif de la recherche qui visait à documenter le savoir de terrain des enseignantes, cette analyse a produit une cartographie du territoire symbolique de l'évaluation formative en termes de production scientifique, territoire divisé en trois « zones de savoir » en tant que « manières de faire » (figure 1).

Figure 1 : Territoire symbolique de l'évaluation formative



En effet, à la suite des travaux de Becker (1982/2006), les « manières de faire partagées » ont été identifiées comme étant des conventions de la culture

socioprofessionnelle des enseignantes, soit des savoirs pratiques qui, s'étant révélés fructueux au fil du temps, se sont cristallisés en accords mutuels tacites relatifs à des significations devenues typiques pour ce groupe. Comme le souligne Becker, les conventions facilitent la coopération et la coordination entre les membres. De même, selon la théorie de la structuration de Giddens (1987), les « manières de faire admises » ont été associées à des savoir-faire singuliers s'étant exprimés en termes de routines et de « théories-en-usage ». Giddens affirme que les routines constituent des dimensions structurantes privilégiées de l'activité sociale quotidienne ; elles seraient façonnées par la capacité des acteurs à transformer les règles et les normes en ressources pour (re)produire l'action, ainsi que par un ensemble de croyances, issues de valeurs, de stratégies et de conceptions implicites, auxquelles adhèrent les acteurs, soit des théories qui orientent leurs conduites. Enfin, le concept d'« accords pragmatiques » (Dionne & Ouellet, 1990, cités par Pépin, 1994) se rapporte à la zone de savoir associée aux « manières de faire contestées » qui interroge les frontières du territoire. De fait, les enseignantes ont développé certaines pratiques – non usuelles – pour s'accommoder des pressions sociales et des normes institutionnelles en matière d'évaluation, tout en maintenant le cap sur la visée de soutien aux apprentissages de l'évaluation formative. Il s'agit d'un « savoir-y-faire » viable, pourrait-on dire, dans le contexte de reddition de comptes qui se généralise actuellement et que les enseignantes estiment contraignant pour la pratique de l'évaluation formative.

En somme, cette recherche collaborative a conduit à identifier un ensemble de « manières de faire » en tant que modèles qui structurent l'agir quotidien, à partir du sens que l'évaluation formative revêt pour les praticiennes qui la mettent en œuvre. Elle a également permis de théoriser le savoir enseignant en identifiant trois zones – partagée, admise et contestée – d'après les significations coproduites et négociées par le groupe d'enseignantes. L'hypothèse épistémologique défendue ici renvoie à l'idée selon laquelle les négociations sociodiscursives favorisées par les discussions entre pairs seraient porteuses des savoirs utiles aux activités quotidiennes organisées. En d'autres termes, comme le formule Giddens (1987) en se fondant sur les travaux de Wittgenstein, le langage commun véhiculé entre membres d'un groupe partageant les mêmes activités est susceptible de rendre compte de la façon dont est conçue la pratique.

UNE RECHERCHE COLLABORATIVE ÉTUDIANT LES RÉGULATIONS INTERACTIVES ENTRE L'ENSEIGNANT ET SES ÉLÈVES

Voyons maintenant ce qu'il en est de la recherche collaborative menée dans le contexte de l'école primaire genevoise. Celle-ci a étudié les pratiques

d'évaluation formative de neuf enseignants⁵ volontaires du cycle moyen (élèves de 8 à 12 ans). L'objectif était d'appréhender les significations que les enseignants pouvaient élaborer collectivement à propos d'une évaluation foncièrement imbriquée dans les processus d'enseignement. Six rencontres par année ont eu lieu pendant trois ans consécutifs ; chaque rencontre durait une journée complète. Le dispositif est décrit en détail dans Mottier Lopez *et al.* (2010). Nous n'exposons ici que quelques caractéristiques qui sont utiles à notre réflexion épistémologique.

La coélaboration d'un questionnement de recherche sur la régulation interactive

À partir du thème général proposé lors de la prise de contact avec les enseignants, « les pratiques d'évaluation formative et la régulation différenciée des apprentissages », l'idée était de codéfinir avec les praticiens un questionnement de recherche. Très vite, il est apparu que les enseignants s'interrogeaient sur leurs pratiques de *guidage interactif* auprès des élèves dans des tâches complexes. Comment intervenir « à chaud », sans instrumentation particulière, dans une visée de régulation de l'apprentissage de l'élève ? Comment caractériser ce guidage formatif ? Qu'est-ce qui le distingue d'une « simple » interaction ?

Le concept de « régulation interactive » (Allal, 1988), familier dans la culture des enseignants genevois de par leur formation initiale et continue, a servi alors d'unité conceptuelle pour coélaborer le questionnement de recherche, sur la base, initialement, de traces de pratiques librement recueillies par chaque enseignant. Dans le modèle théorique de l'évaluation formative dite « élargie » (Allal & Mottier Lopez, 2005), le concept de régulation désigne les adaptations de l'enseignement et les processus d'ajustement chez l'apprenant qui s'effectuent en exploitant les informations produites par l'évaluation, y compris informelle. La « régulation interactive » a une fonction de médiation en cours de tâche, dans le but de soutenir les progressions d'apprentissage des élèves. Un travail théorique a été réalisé à partir de textes apportés par les enseignants et la chercheure. Enrichi par ce partage intersubjectif de savoirs scientifiques et de pratiques rapportées, le questionnement de recherche a progressivement émergé et a été formulé en ces termes à la fin de la première année : *Dans quelle mesure certaines interactions entre l'enseignant et ses élèves relèvent-elles d'une régulation interactive à des fins d'évaluation formative ? Dans des situations complexes de résolution de problèmes mathématiques, quelle est la nature du guidage de l'enseignant qui vise à soutenir les activités et les apprentissages des élèves ? Quelles sont les attentes et les significations que les enseignants attribuent aux régulations formatives entreprises ?*

5. Ils ont de 7 à 33 ans d'expérience professionnelle.

Pour y répondre, deux sources de données principales ont été décidées collectivement : 1) un enregistrement audio de toutes les réunions de recherche et un recueil des traces écrites produites à ces occasions ; 2) un enregistrement audio d'interactions sociales réalisé par chaque enseignant dans sa classe et un recueil des traces écrites des élèves s'y rapportant. Tout comme dans la première recherche présentée plus haut, les enseignants ont sélectionné eux-mêmes, en dehors des réunions, des extraits d'interactions qu'ils jugeaient représentatifs d'une régulation interactive à des fins d'évaluation formative. Avant la réunion, un « fichet » associé à chaque extrait transcrit par l'enseignant devait être rempli, selon un ensemble de catégories concernant notamment les attentes de l'enseignant, le déroulement didactique de la leçon et les caractéristiques de l'élève concerné. Une écriture réflexive a aussi été sollicitée sur l'intérêt de l'épisode du point de vue de l'enseignant. Les fichets ont été partagés en réunion. Ce recueil de données dans les classes a eu lieu au cours de la deuxième année de la recherche, puis une seconde fois lors de la troisième année.

Des analyses collectives des extraits d'interactions choisis par les enseignants

De par sa conception, la recherche collaborative menée visait à construire au sein du groupe une culture de partage de responsabilités et d'expertises de différents types au service du projet collectif (St-Arnaud, 2003). La chercheuse offrait ainsi au groupe des ressources d'un certain ordre (savoirs théoriques sur la régulation en classe et savoirs méthodologiques) ; de même, les enseignants proposaient des ressources qui leur étaient propres (connaissances du contexte, expériences professionnelles, savoirs pratiques).

Les données issues des réunions ont, pour leur part, été examinées par la chercheuse. Celle-ci a effectué au fur et à mesure des analyses de contenu des échanges intersubjectifs, puis a proposé des synthèses soumises à l'interprétation des enseignants – à des fins de validité de signifiante – à nouveau objectivées par des comptes rendus, dans un processus circulaire proche de celui présenté dans la première recherche.

Pour les données récoltées dans les classes, l'objectif était d'impliquer les enseignants dans des démarches de coanalyse et de co-interprétation des épisodes interactifs. Le choix opéré a été d'introduire auprès des enseignants deux outils d'analyse issus de propositions de la littérature scientifique. Ce choix répondait par ailleurs à des demandes explicites de leur part à la chercheuse : « On attend de toi que tu nous apportes ce qui existe déjà » ; « On a besoin d'outils que tu possèdes, on ne veut pas tout réinventer. »

Le premier outil proposait un cadre d'analyse du guidage interactif de l'enseignant développé à partir de travaux en linguistique (Pekarek, 1999) et adapté au contexte scolaire (Mottier Lopez, 2008). Les principaux paramètres discursifs définis dans ce cadre d'analyse permettaient de caractériser la régulation interactive : propriétés d'un guidage « ciblé » (explication de contenus par l'enseignant, sollicitation de l'élève concernant des éléments présents dans l'environnement de la classe, sollicitation de réponses dites de restitution) *versus* propriétés d'un guidage « ouvert » (sollicitation de réponses dites de développement, sollicitation d'échanges entre élèves, d'initiatives de leur part, etc.). Le deuxième outil visait à identifier les opérations de régulation en acte, à partir des propositions de Vermunt et Verloop (1999) également ancrées dans le contexte scolaire : anticipation, contrôle en cours d'action, ajustement en cours d'action, évaluation en arrêt ou fin de tâche.

Ces cadres préexistants, à visée *descriptive*, ont permis des analyses systématiques et ont instrumenté les échanges intersubjectifs entre les participants. Les constats ont été confrontés aux référents des enseignants par rapport à leurs attentes et aux contraintes situées qu'ils perçoivent à la régulation des apprentissages de leurs élèves : programme à parcourir et échéances des évaluations certificatives, temps didactique, gestion du groupe-classe et des différences entre élèves, cohérence avec les principes des moyens d'enseignement qui, par exemple, privilégient la dévolution aux élèves des problèmes à résoudre.

Quelques résultats produits sur les guidages formatifs des enseignants

Deux plans ont fait l'objet de résultats produits avec les enseignants. Le premier porte sur les analyses des épisodes confrontés aux contenus des fichets (écriture réflexive subjective) et aux échanges intersubjectifs pendant les réunions. Le deuxième concerne le développement professionnel des enseignants. Dans ce chapitre, nous nous centrerons sur quelques constats formulés collectivement à propos des pratiques observées en classe.

Des variations de guidage formatif ont été observées chez un même enseignant en fonction des contributions interactives de ses élèves. En cas de difficulté dans la résolution du problème, l'enseignant a tendance, par son intervention, à transformer la situation complexe en pointant au cours de l'interaction les connaissances et les ressources nécessaires pour réaliser la tâche. Les enseignants se disent toutefois *insatisfaits* lorsqu'ils se sentent contraints de procéder à un tel guidage (souvent sous forme de maïeutique), car ils préfèrent une dynamique interactive propice à des réponses d'explicitation de raisonnements et de démarches « originales »

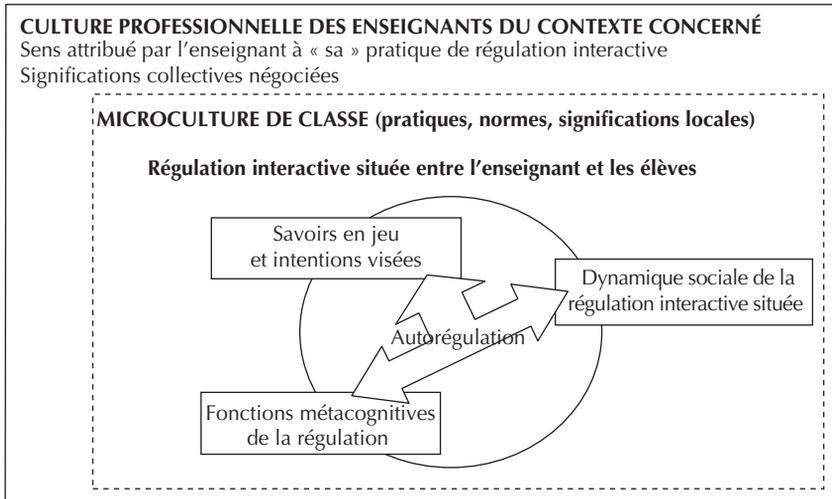
non préstructurées par un questionnement ciblé. L'analyse de l'ensemble des extraits recueillis dans les classes a montré que le guidage ciblé dans la régulation interactive se manifeste tendanciellement avec des élèves de niveau scolaire faible en mathématiques. Les enseignants disent se sentir démunis : « On ne sait pas comment faire autrement avec ces élèves... sinon ils décrochent dans les résolutions de problèmes. » Par rapport aux principes didactiques qui privilégient la dévotion des problèmes à l'élève, les enseignants disent ne pas se sentir autorisés à soutenir l'activité de l'élève par un tel guidage. De plus, les analyses des épisodes tendent parfois à montrer des réponses de conformité de la part de l'élève qui ne témoignent pas de la construction d'une réelle compréhension commune dans la régulation formative. Autrement dit, un guidage ciblé ne crée pas forcément les effets attendus par l'enseignant, mais permet à l'élève de se réinvestir dans la tâche. La régulation porte alors plus sur l'activité de l'élève que sur son raisonnement mathématique. Au fil des analyses collectives et des co-interprétations, les enseignants se sont accordés sur le fait que quel que soit le type de guidage à des fins d'évaluation formative, il importait que leurs interventions soient en mesure de s'ajuster au mieux au raisonnement de l'élève, et qu'ils soient eux-mêmes davantage conscients des enjeux d'apprentissage et des variables didactiques pour penser l'intervention formative.

Comme cette brève description le montre, les outils d'analyse pré-existants ont proposé des catégories conceptuelles que le groupe s'est approprié pour décrire les pratiques étudiées. Ils constituent des outils symboliques qui offrent une médiation à la réflexion critique des enseignants et de la chercheuse sur les pratiques situées. C'est le cas, en particulier, pour un aspect informel et difficile à appréhender de l'évaluation formative : la régulation interactive qui a pour caractéristique de se « fondre » dans les activités conjointes entre l'enseignant et ses élèves. L'utilisation de ces outils conceptuels a permis de reconnaître collectivement des propriétés à la régulation interactive, à partir desquelles des confrontations de points de vue et des négociations de sens ont pu se développer, ouvertes aux savoirs pratiques des enseignants.

Construction d'un modèle à partir de la recherche collaborative

Outre les résultats présentés plus haut sur les pratiques contextualisées des enseignants, la recherche a offert les conditions à l'élaboration d'un modèle de la régulation interactive dans une perspective située. La figure 2 expose ce modèle qui est par ailleurs détaillé dans Mottier Lopez (2012). Il présente une amorce de réorganisation conceptuelle par rapport à des savoirs existants pour étudier les régulations interactives en classe.

Figure 2 : Modèle d'analyse de la régulation interactive entre l'enseignant et ses élèves



Commençons par les catégories d'analyse, vues comme interreliées, de la régulation interactive entre l'enseignant et les élèves qui, dans une perspective d'évaluation formative, vise à soutenir l'autorégulation de l'élève ; cette autorégulation est elle-même nécessaire au développement de la régulation interactive située :

- Les savoirs en jeu et les intentions visées par la régulation interactive au regard des pratiques et normes de la microculture de classe (Mottier Lopez, 2008). Par le moyen d'analyses a priori et a posteriori sont identifiés notamment les objectifs visés, types de tâche, contraintes et ressources des situations, erreurs et difficultés des élèves (anticipées et observées) ;
- La dynamique interactive de la régulation, en examinant la nature du guidage interactif de l'enseignant et les attentes que ce dernier a par rapport à la participation interactive de ses élèves. Il s'agit ici d'appréhender la structure sociale de la régulation interactive ;
- Les opérations de régulation telles qu'elles sont classiquement définies dans la littérature, afin notamment de constater à quel moment précis la régulation interactive se déploie en cours de réalisation de tâche. Ce faisant, ce sont les fonctions métacognitives de la régulation qui sont appréhendées : anticipation, contrôle, ajustement, évaluation.

La régulation interactive suppose aussi une interaction avec les produits de la culture, outils culturels et symboliques de médiation qui participent pleinement à l'autorégulation des conduites individuelles. Sans les développer ici, cela nous amène à prendre en considération deux niveaux

interreliés⁶ : la microculture de classe, appréhendée en termes de pratiques, normes, significations négociées dans le contexte local de chaque classe (Mottier Lopez, 2008) et la culture professionnelle des enseignants. Dans notre recherche collaborative, nous avons considéré :

- Le sens situé que chaque enseignant construit par rapport à sa propre pratique de régulation interactive, en fonction de la connaissance qu’il possède de son contexte d’exercice professionnel, de ses élèves, de la culture de son établissement scolaire, en fonction de ses propres valeurs, etc. ;
- Les significations négociées collectivement dans le groupe social impliqué, ouvrant à de nouvelles compréhensions possibles et à la construction d’une culture partagée de l’évaluation formative au regard aussi du contexte socioprofessionnel des enseignants.

En somme, ce modèle prend en compte les propriétés de ce que représente, d’un point de vue théorique, une régulation interactive formative, tout en conceptualisant ce qui fait sens pour les enseignants sur un plan individuel et à la fois collectif. L’hypothèse épistémologique est que, ce faisant, un « savoir » sur la régulation interactive se construit dans un rapport *dialectique* (au sens de Lave, 1988) entre des savoirs scientifiques, formalisés dans les outils d’analyse utilisés avec les enseignants, et les savoirs d’expérience, objectivés à travers les négociations de sens dans le groupe social que représente la recherche collaborative.

LA RECHERCHE COLLABORATIVE : QUELS ÉCLAIRAGES SUR L’ÉVALUATION FORMATIVE ?

Notre projet consistait à présenter l’architecture de deux recherches collaboratives portant sur l’évaluation formative en salle de classe, afin de mieux comprendre comment leurs présupposés théoriques et épistémologiques, de même que les choix méthodologiques, affectent les modèles impliqués et les savoirs produits dans ce type de recherche.

Ainsi, nous avons montré que s’inspirer de la « tradition de Chicago », comme le fait la première recherche, c’est tenter de décrire au mieux ce que font les acteurs de terrain ; c’est privilégier un regard de l’intérieur sur l’objet d’étude. Cette ouverture est en soi une prise de position par rapport au savoir à coconstruire, tributaire de la posture compréhensive de la chercheuse et de son cadre théorique a priori qui repose sur un modèle d’acteur compétent (Giddens, 1987) et sur des travaux associés au domaine des savoirs pratiques (Schön, 1983/1994). Cette posture oriente le déroulement de l’investigation en suggérant que les enseignantes s’engagent, accompagnées par la chercheuse, à expliciter leurs « manières de faire »,

6. Sachant par ailleurs que d’autres niveaux de contexte culturel existent que nous ne pouvons développer dans le cadre de ce chapitre, telle la culture de l’établissement scolaire.

à en négocier la validité entre elles, et à en offrir une théorisation plus ou moins formelle. En ce sens, les types de savoirs qui émergent de cette complémentarité des rôles sont, entre autres, des modèles de la mise en œuvre de l'évaluation formative. Ces modèles qui organisent la pratique sont des savoirs objectivés : d'abord par les pairs, qui ont aussi une compréhension de la mise en œuvre en contexte de l'évaluation formative et qui sont donc considérés comme les mieux à même d'en juger la viabilité et la légitimité selon l'option théorique de l'acteur compétent. Puis ils le sont par une mise en relation avec d'autres cadres théoriques dont la pertinence émerge lors de l'analyse a posteriori. Ces cadres viennent insuffler un sens nouveau au matériau recueilli lors de la démarche collaborative en questionnant et en situant les nouveaux savoirs parmi le corpus de recherches existant.

Par ailleurs, nous avons montré avec la deuxième recherche que s'inspirer d'une approche inductive-déductive (Balslev & Saada-Robert, 2002) qui tente de penser ensemble le préexistant et l'émergent (Lave, 1988), c'est mettre à la disposition des participants de la recherche des savoirs scientifiques constitués et validés pour nourrir la réflexion sur les pratiques objectivées et confrontées dans l'intersubjectivité. Le but est d'offrir des outils de médiation aux analyses collectives en tentant le pari d'un métissage entre des composantes scientifiques, praxiques et sociales par le moyen de l'objectivation critique. Dans ce cadre, la dynamique des rôles distribués qui s'établit entre le chercheur et les praticiens renvoie à une autre idée de la collaboration, davantage interventionniste au plan de la conceptualisation de l'objet étudié et de la validation négociée des savoirs produits. Le modèle élaboré à partir de cette recherche est d'ordre théorique, en tant que cadre d'analyse et d'interprétation d'une pratique difficile à appréhender : des démarches interactives et informelles d'évaluation formative intégrées dans les processus d'enseignement. Le savoir qui en résulte est local par certains aspects et, pour d'autres aspects, susceptible d'être comparé à d'autres résultats de recherche en raison des unités conceptuelles dont le sens est reconnu au-delà du groupe même de la recherche collaborative.

Par delà les différences d'architecture des démarches collaboratives mises en œuvre et de la façon dont les chercheuses ont usé des cadres théoriques, les deux recherches accordent une importance capitale à l'intégration du point de vue des praticiens – leur réflexivité pratique – dans la production de savoirs à propos de l'objet de préoccupation mutuelle. La production scientifique repose ainsi sur une théorisation du savoir de terrain, au profit d'une meilleure compréhension d'une pratique étroitement liée à ce qui est tenu pour légitime par les praticiens dans le cadre de leurs repères socioprofessionnels. Nous pensons que c'est une des conditions pour que la recherche scientifique soit en mesure de produire des savoirs « en résonance » avec les pratiques. Une autre condition est de prendre très sérieusement en compte le système socioculturel et historique plus vaste dans lequel les pratiques d'évaluation formative s'inscrivent et auquel elles contribuent. Un enjeu scientifique est

d'appréhender la culture de métier dont elles relèvent et qui requiert une validation par l'intersubjectivité, le collectif, le socialement négocié – y compris à travers la confrontation aux savoirs scientifiques constitués. Ces « savoirs négociés » sont susceptibles, selon nous, d'offrir des contributions significatives au domaine de l'évaluation des apprentissages des élèves en classe.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. (1988). Vers un élargissement de la pédagogie de la maîtrise : processus de régulation interactive, rétroactive et proactive. In M. Huberman (Éd.), *Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise* (pp. 86-126). Neuchâtel : Delachaux et Niestlé.
- Allal, L. & Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. In Organisation de coopération et de développement économiques (Éd.), *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires* (pp. 265-290). Paris : OCDE.
- Balslev, K. & Saada-Robert, M. (2002). Expliquer l'apprentissage de la littéracie. Une démarche inductive/déductive. In F. Leutenegger & M. Saada-Robert (Éd.), *Expliquer et comprendre en sciences de l'éducation* (pp. 89-110). Bruxelles : De Boeck.
- Becker, H. S. (1982/2006). *Les mondes de l'art* (2^e éd.). Paris : Flammarion.
- Black, P. & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Éd.), *Assessment and learning* (pp. 9-25). London : Sage.
- Bourassa, M., Bélair, L. & Chevalier, J. (2007). Les outils de la recherche participative. *Éducation et francophonie*, 35(2), 1-11.
- Chapoulie, J.-M. (2001). *La tradition sociologique de Chicago. 1892-1961*. Paris : Seuil.
- Dann, R. (2002). *Promoting assessment as learning. Improving the learning process*. London : Routledge and the Falmer Press.
- De Certeau, M. (1990). *L'invention du quotidien* (vol. 1). Paris : Gallimard.
- Desgagné, S. (2001). La recherche collaborative : nouvelle dynamique de recherche en éducation. In M. Anadon (Éd.), *Nouvelles dynamiques de recherche en éducation* (pp. 51-76). Québec : Presses de l'Université Laval.
- Dionne, P. & Ouellet, G. (1990). *La communication interpersonnelle et organisationnelle : l'effet Palo Alto*. Paris : Éditions d'Organisation/Boucherville, Québec : Gaëtan Morin.
- Giddens, A. (1987). *La constitution de la société*. Paris : Presses Universitaires de France.
- Lave, J. (1988). *Cognition in practice : Mind, mathematics and culture in everyday life*. Cambridge, MA : Cambridge University Press.
- Morrisette, J. (2009). La portée d'une perspective socioculturelle de l'évaluation formative : vers l'élargissement d'une conceptualisation. *Mesure et évaluation en éducation*, 32(2), 1-23.

- Morrisette, J. (2010a). *Manières de faire l'évaluation formative des apprentissages : analyse interactionniste du savoir-faire d'enseignantes du primaire*. Sarrebruck, Suisse : Les Éditions universitaires européennes.
- Morrisette, J. (2010b). Un panorama de la recherche sur l'évaluation formative des apprentissages. *Mesure et évaluation en éducation*, 33(3), 1-27.
- Morrisette, J. (2011). Vers un cadre d'analyse interactionniste des pratiques professionnelles. *Recherches qualitatives*, 30(1), 38-59. Consulté le 15 janvier 2012 dans http://www.recherche-qualitative.qc.ca/numero30%281%29/RQ_30%281%29_Morrisette.pdf
- Mottier Lopez, L. (2006). L'évaluation située des apprentissages : le rôle fondamental de la médiation sociale. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 82-89). Paris : L'Harmattan.
- Mottier Lopez, L. (2008). *Apprentissage situé : la microculture de classe en mathématiques*. Berne : Peter Lang.
- Mottier Lopez, L. (2012). *La régulation des apprentissages en classe*. Bruxelles : De Boeck.
- Mottier Lopez, L., Borloz, S., Grimm, K., Gros, B., Herbert, C., Methenitis, J., Payot, C., Pot, M., Schneuwly, J.-L. & Zbinden, S. (2010). Les interactions de la régulation entre l'enseignant et ses élèves, expérience d'une recherche collaborative. In L. Lafortune, S. Fréchette, N. Sorin, P.-A. Doudin & O. Albanese (Éd.), *Approches affectives, métacognitives et cognitives de la compréhension* (pp. 33-50). Québec : Presses de l'Université du Québec.
- Pekarek, S. (1999). *Leçons de conversation. Dynamiques de l'interaction et acquisitions de compétences discursives en classe de langue seconde*. Fribourg : Éditions universitaires.
- Pépin, Y. (1994). Savoirs pratiques et savoirs scolaires : une représentation constructiviste de l'éducation. *Revue des sciences de l'éducation*, 20(1), 63-85.
- Schön, D. A. (1983/1994). *Le praticien réflexif. À la recherche du savoir caché dans l'agir professionnel*. Montréal : Logiques.
- St-Arnaud, Y. (2003). *L'interaction professionnelle : efficacité et coopération*. Montréal : Les Presses de l'Université de Montréal.
- Tochon, F. (1996). Rappel stimulé, objectivation clinique, réflexion partagée. Fondements méthodologiques et applications pratiques de la rétroaction vidéo en recherche et en formation. *Revue des sciences de l'éducation*, 22(3), 467-502.
- Van Campenhout, L., Chaumont, J.-M. & Franssen, A. (2005). *La méthode d'analyse en groupe. Application aux phénomènes sociaux*. Paris : Dunod.
- Vermunt, J. D. & Verloop, N. (1999). Congruence and Friction between Learning and Teaching. *Learning and Instruction*, 9(3), 257-280.

Entre savoirs scientifiques et usages normatifs : les enquêtes internationales d'évaluation des acquis des élèves

**Dominique Lafontaine & Christian Monseur
Université de Liège**

INTRODUCTION

Depuis le lancement du Programme international pour le suivi des acquis des élèves (PISA) par l'Organisation de coopération et de développement économiques (OCDE) en 2000, les résultats des enquêtes internationales sont au cœur du débat médiatique et leur retentissement réel ou symbolique sur les politiques éducatives est loin d'être négligeable. Les résultats des évaluations ainsi diffusés dans le grand public y jouissent d'une aura de scientificité qui leur confère souvent un caractère absolu ou universel, dépassant de loin la mesure avec laquelle les scientifiques producteurs de ces savoirs conseilleraient de les interpréter.

Le questionnement épistémologique s'orientera dans deux directions : dans un premier temps, nous nous attacherons à montrer, en utilisant les données de PISA et d'IEA-PIRLS¹, en quoi des choix théoriques – relatifs au domaine évalué ou à la mesure – ou méthodologiques (définition de la population de référence, choix des textes) faits en amont contraignent la nature et les résultats de l'enquête et en limitent la portée. Dans un second temps, notre attention se portera sur l'aval, en l'occurrence les usages des savoirs produits par les enquêtes internationales, en particulier

1. *International Association for the Evaluation of Educational Achievement (IEA) – Progress in International Reading Literacy Study (PIRLS).*

PISA². On s'interrogera sur la nature et la forme des savoirs produits et sur l'usage qui en a été fait par les différents acteurs de la scène éducative, principalement au cours de la dernière décennie³. À la lumière de leur propre positionnement, les auteurs tenteront enfin, en guise de conclusion, de mettre en perspective la posture singulière du chercheur spécialiste de l'évaluation dans ce type d'enquête internationale, à la croisée des mondes scientifique et politique.

À cet égard, il importe de jouer cartes sur table : les deux auteurs de cette contribution ont assumé, et assument encore, des fonctions sur le plan national et international dans les enquêtes PISA et les enquêtes de l'IEA⁴. L'éclairage qu'ils apporteront sur les enquêtes internationales porte la marque de leur positionnement institutionnel. Il s'agit donc d'un exercice d'analyse critique venu de l'intérieur ; s'il présente le risque d'une forme de parti pris⁵, il présente aussi l'avantage de s'appuyer sur la solide connaissance des enquêtes internationales et des contextes internationaux qu'en ont les auteurs.

LES THÉORIES AU FONDAMENT DES ENQUÊTES INTERNATIONALES

Les grandes enquêtes internationales contemporaines comme PISA, PIRLS ou TIMSS⁶ ne bâtissent pas sur du sable, contrairement à ce qu'en pensent parfois le public non averti et un public qui devrait l'être⁷. Des théories sont en effet mobilisées pour l'élaboration et la mise en œuvre de ces enquêtes.

2. Le choix de cette enquête se justifie d'une part parce que c'est l'enquête internationale dont la stratégie et les ambitions d'influencer les politiques éducatives sont le plus clairement affirmées ; d'autre part, parce que c'est aussi une enquête pour laquelle davantage d'informations sont disponibles sur l'influence que peuvent avoir ces enquêtes dans les différents systèmes éducatifs (Grek, 2009).

3. L'évolution qui consiste à faire des enquêtes internationales des acquis des élèves un instrument de gouvernance est récente et a vraisemblablement été précipitée par l'arrivée de PISA. Les précédentes enquêtes, menées à l'initiative de l'IEA depuis la fin des années 1960 n'avaient pas, du moins avant le milieu des années 1990, le même retentissement en dehors du champ scientifique.

4. Dominique Lafontaine est membre du Groupe d'experts en lecture PISA depuis 1999 et a été membre du Groupe d'experts en lecture pour IEA-PIRLS 2006 ; elle a été à plusieurs reprises gestionnaire nationale pour les études de l'IEA (1991) et PISA (2000, 2003). Christian Monseur est membre du Groupe de conseil technique (*Technical Advisory Group*) de PISA et de l'IEA ; il a été gestionnaire national pour TIMSS 1995.

5. Risque que n'évitent d'ailleurs pas les critiques venues de l'extérieur.

6. *Trends in International Mathematics and Science Study* (TIMSS).

7. Un professeur d'université spécialiste des sciences s'exprimant dans la presse avait commenté PISA en ces termes : « PISA est un questionnaire Marie-Claire ou Nouvel Obs' d'été, qui n'a aucune valeur » (*Le Soir*, 22 mars 2009, p. 24), comparaison qui en disait long sur son degré d'information...

Elles sont principalement de deux ordres : des théories relatives au domaine évalué (la compréhension en lecture, les mathématiques, les sciences) et des théories relatives à la mesure.

Des théories relatives au domaine évalué : le cadre de référence

De nos jours, préalablement à toute épreuve internationale, un cadre d'évaluation (*framework*) est défini (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009 ; Mullis *et al.*, 2005 ; OCDE, 2009b). Il n'en a pas toujours été ainsi. Les études internationales les plus anciennes, menées sous les auspices de l'IEA⁸, entre les années 1968 et 1995, ne comportaient pas de cadre de référence formalisé⁹. À cet égard, les enquêtes TIMSS (Robitaille *et al.*, 1993) et PISA (OCDE, 1999) marquent un véritable tournant. Depuis lors, toute épreuve internationale comporte ce document « fondateur » qu'est le cadre de référence pour l'évaluation. Il revient généralement à un groupe ad hoc d'experts internationaux du domaine évalué de produire une première version de ce cadre de référence ; celle-ci est largement mise en débat auprès des représentants des systèmes éducatifs ou des gestionnaires nationaux de l'enquête, ensuite révisée afin d'arriver à une version définitive qui puisse recueillir le consensus des pays participant à l'étude.

Typiquement, un cadre de référence envisage les aspects suivants :

- Quel est l'objet de l'évaluation ? Quelle définition se donne-t-on de l'objet ? À quels modèles théoriques reconnus dans la littérature fait-on référence pour évaluer ce domaine ?
- Comment le domaine que l'on veut évaluer se structure-t-il ? Quels aspects, quelles compétences, quels savoirs et contenus l'évaluation doit-elle prendre en compte ?
- Quelles seront les modalités d'évaluation utilisées ?¹⁰
- Comment présenter et interpréter les résultats de l'évaluation ? Considère-t-on qu'il existe une compétence unique représentée par un score global sur une échelle, ou les compétences sont-elles multidimensionnelles, appelant plusieurs scores distincts, plusieurs sous-échelles ?
- Quelles dimensions envisager dans les questionnaires contextuels, destinés à l'élève, au chef d'établissement ? Sur quelles théories sociologiques et pédagogiques ces questionnaires s'appuient-ils ?

8. Voir <http://www.iea.nl>

9. Elles comportaient toutefois une analyse approfondie des curriculums nationaux (voir par exemple, Rosier & Keeves, 1991 ; Travers & Westbury, 1989). Les études IEA continuent à produire cette étude du curriculum (par exemple pour PIRLS 2006, Kennedy, Mullis, Martin & Trong, 2007).

10. Le cadre de référence, typiquement, passe en revue tous les facteurs qui pourraient influencer la difficulté des items.

Pour prendre l'exemple de la lecture dans PISA, le cadre de référence élaboré pour PISA 2000 a été considérablement révisé en 2009 (cycles où la lecture était le domaine majeur) pour incorporer les évolutions intervenues dans le domaine au cours des 10 années précédentes, en particulier la composante lecture de textes électroniques (OCDE, 2009b). Dans quelle mesure la lecture en ligne diffère-t-elle de la lecture classique sur papier ? Fait-elle appel à d'autres compétences, d'autres habiletés ? En quoi les textes électroniques diffèrent-ils des textes papier ? Comment développer des tâches de lecture électronique suffisamment ouvertes pour évaluer les compétences et habiletés visées (par exemple sélectionner l'URL la plus pertinente pour une recherche d'information, évaluer la fiabilité d'une source sur internet) sans ouvrir un univers infini comme l'est internet – ce qui poserait d'insurmontables problèmes de traduction ? Toutes ces questions sont abordées, discutées en profondeur et trouvent une réponse dans le cadre de référence.

Ces documents, précieux pour appréhender les théories au fondement des enquêtes internationales, aisément accessibles auprès des organismes responsables de ces enquêtes (IEA, OCDE), sont trop souvent méconnus non seulement du grand public, mais aussi des détracteurs des enquêtes internationales qui font peu de cas des assises théoriques du domaine évalué, telles la nature et la structure des textes dans le domaine de la lecture, pour ne prendre qu'un exemple.

Des théories relatives à la mesure

Si les théories relatives au domaine évalué sont aisément accessibles, il n'en va pas de même des théories relatives à la mesure. Certes, tous les détails techniques compréhensibles par un public familier de ces aspects sont consignés dans un copieux rapport technique (Martin, Mullis & Kennedy, 2007 ; OCDE, 2009a). Mais les théories au fondement même de la mesure font rarement l'objet d'une explicitation dans les documents qui accompagnent les enquêtes internationales ; elles font partie de ces évidences partagées dans un paradigme de recherche qui sont rarement mises à plat de l'intérieur du paradigme, mais plus souvent dénoncées ou critiquées par des tenants d'autres paradigmes. Ces évidences sont nombreuses et varient fortement dans leur degré de généralité. Essayons d'en dégager quelques-unes. Ainsi, en allant du plus général au plus particulier¹¹, les enquêtes internationales :

- Partent du principe que l'on peut mesurer d'une manière fiable et valide les savoirs, savoir-faire et connaissances (*knowledge and skills*) des élèves par un ou des tests standardisés – ce que d'aucuns contestent,

11. Cet inventaire est loin d'être exhaustif.

notamment les tenants de la cognition située qui considèrent que les savoirs sont enracinés dans les contextes et situations¹² ;

- Considèrent que l'on peut mesurer ces dimensions d'une manière fiable et relativement juste dans un contexte international ; mettent dès lors en œuvre un arsenal sophistiqué de procédures de standardisation et de vérification pour s'assurer de la comparabilité des données, de l'équivalence des épreuves sur le plan linguistique en particulier ;
- S'inscrivent dans une tradition de recherche éducatrice qui accorde du crédit à des dimensions telles que la validité, la fidélité (et en conséquence aux indices qui évaluent la qualité des tests à cet égard), voire qui donne le pas à ces dimensions par rapport à la notion de pertinence¹³ ;
- Considèrent que la performance ou compétence d'un individu¹⁴ dans un domaine peut notamment se résumer à un score unique sur une échelle (ou des scores sur quelques sous-échelles) ; que l'on peut donc « sommer » les réponses données à des items et en inférer quelque chose qui représente la compétence ou la performance d'un individu dans un domaine ;
- Produisent des estimations de la performance ou compétence des sujets testés qui sont vraiment très éloignées de l'estimation classiquement utilisée à l'école pour juger de la performance, soit le pourcentage de réponses correctes à une épreuve généralement identique pour l'ensemble des élèves de la classe. Ce dernier point appelle des développements spécifiques concernant les modèles de la réponse à l'item communément utilisés dans les enquêtes internationales depuis le début des années 1990 (Lafontaine & Raïche, 2011) et qui font l'objet de la section suivante.

12. Pour une réflexion plus approfondie sur approche située et standards internationaux, voir Lafontaine (2010).

13. Selon De Ketele et Gérard (2005), « les grandes enquêtes internationales du type PISA ou TIMMS [sic] [...] sont fiables et valides, mais ne sont pas pertinentes par rapport à l'approche par compétences et donc par rapport aux finalités des systèmes éducatifs conçus dans cet esprit » (p. 5). L'affirmation selon laquelle une enquête comme PISA n'évalue pas de compétences prête toutefois à discussion, d'autres auteurs (Cattotar, Mangez, Delvaux, Mangez & Maroy, 2009) considérant au contraire comme une évidence que PISA s'inscrit dans le vaste mouvement des compétences. Par ailleurs, contester sa pertinence au motif qu'elle ne correspondrait pas aux finalités des systèmes éducatifs qui prônent une approche par compétences nous paraît aller vite en besogne. Si on prend le cas de la Communauté française de Belgique, Demonty et Fagnant (2004) ont montré que dans le domaine des mathématiques, les épreuves PISA correspondent assez fidèlement aux « compétences » figurant dans les référentiels communs à l'ensemble des établissements (document qui correspond aux finalités officielles de l'enseignement). Si en outre on s'appuie sur la distinction classique entre curriculum visé et curriculum implanté, on pourrait même soutenir que les épreuves PISA sont bien plus proches d'une évaluation de compétences que ne le sont la majorité des épreuves d'évaluation pratiquées en salle de classe.

14. Simple traduction du terme anglais *proficiency* qui ne règle pas l'épineuse question débattue surtout dans le monde francophone de savoir si les enquêtes internationales mesurent ou non des compétences.

LES MODÈLES DE LA RÉPONSE À L'ITEM : FONDEMENTS ET IMPLICATIONS

L'utilisation des modèles de la réponse à l'item dans les enquêtes internationales a un impact qui dépasse de loin les possibilités techniques qu'autorisent ces modèles. Concrètement, c'est grâce à ces modélisations que l'on peut mettre en œuvre des plans d'évaluation complexes avec rotation de livrets, ce qui permet d'évaluer en profondeur le domaine sans surcharger chacun des élèves testés. C'est aussi grâce à ces modélisations que l'on peut mesurer de façon rigoureuse les évolutions dans le temps (indicateurs de tendances) ou encore décrire, d'une manière précise, les compétences d'élèves situés à différents niveaux sur les échelles de résultats. Mais de quoi s'agit-il exactement ? Au risque d'une simplification, nous allons tenter ci-dessous d'en faire comprendre le principe et les implications.

Le principe de base des modèles de la réponse à l'item est qu'ils permettent d'exprimer et de mettre en correspondance, sur une même échelle, le niveau de difficulté d'un item et le niveau de compétence d'un individu. Un individu qui obtient par exemple un score de 550 sur une telle échelle a une probabilité supérieure à 50 %¹⁵ de réussir les items d'une difficulté inférieure à 550, et une probabilité inférieure à 50 % de réussir les items d'une difficulté supérieure à 550. Ces échelles permettent d'aller au-delà d'une perspective strictement normative consistant à ordonner les individus ou les pays, dans les enquêtes internationales, du plus performant au moins performant. Avec ces modèles, l'information se précise et devient plus diagnostique. Non seulement tel pays sait qu'il est plus ou moins performant que tel autre, mais il sait aussi en quoi : face à quels types de textes par exemple, quelle proportion des individus évalués sont classés à tel ou tel « niveau » et se sont révélés capables de mettre en œuvre telle démarche de lecture ou d'effectuer telle tâche. Ces modèles réussissent en quelque sorte à combiner une approche normative (qui reste au cœur de toute étude comparative) et une approche plus diagnostique. On se plaira à souligner, même si certains contestent le fait que les enquêtes internationales évaluent des compétences (au sens de tâches complexes et inédites)¹⁶, à quel point le continuum hiérarchique des niveaux établis grâce aux modèles de la réponse à l'item épouse la perspective d'une compétence qui se construit pas à pas, dans une progression en spirale.

Dans toutes les enquêtes à grande échelle contemporaines, les individus obtiennent non pas un score unique, mais cinq valeurs plausibles pour chaque échelle ou sous-échelle. On tente de pallier le nombre restreint d'items administrés à chacun des répondants par l'utilisation d'informations

15. Ce seuil est celui généralement utilisé, mais il est modulable dans le programme.

16. « Sur la base d'une analyse de différentes épreuves internationales comme PISA et le PASEC, nous constatons que les performances évaluées se cantonnent aux ressources » (De Ketele, 2010, p. 40).

collatérales : statut socio-économique, sexe, langue parlée à la maison, grade fréquenté, motivation, etc. Les valeurs plausibles sont ainsi produites en fonction non seulement des réponses effectives aux items auxquels l'individu testé a effectivement répondu, mais aussi de ces informations collatérales sur la base desquelles on infère ce que les individus auraient pu répondre à des questions qui ne leur ont pas été administrées¹⁷.

LES PARAMÈTRES MÉTHODOLOGIQUES INFLUENTS

Outre les théories qui président en amont aux grandes orientations des enquêtes internationales, différents choix d'ordre méthodologique ont une influence non négligeable sur la nature et la valeur des indicateurs de résultats issus de ces enquêtes. Nous avons choisi d'illustrer cet impact des choix d'ordre méthodologique sur les indicateurs de performances (moyennes et dispersion), d'équité et de tendances (progrès dans le temps) à partir d'analyses de données issues de PISA.

Définition de la population

Un des premiers choix méthodologiques posé dans le cadre d'une enquête internationale est celui de la population de référence. Plusieurs possibilités existent à cet égard, dont les deux principales sont :

- Une population définie exclusivement par son âge (population dite à âge constant) ; c'est le cas de PISA, qui s'intéresse aux élèves âgés de 15 ans, sans prendre en considération l'année d'études qu'ils fréquentent ; en Communauté française de Belgique, par exemple, on trouve dans l'échantillon des élèves de la 1^{re} à la 6^e année du secondaire, la majorité étant en 4^e secondaire, année que fréquentent les élèves à l'heure dans leur parcours scolaire ;
- Une population définie principalement par son niveau d'études (population dite à niveau d'études ou grade constant), avec des indications ou limites relatives à l'âge des élèves, variables selon les études (il en existe de nombreuses variantes) ; dans PIRLS par exemple, la population de référence est définie comme le grade où se trouvent les élèves ayant reçu quatre années d'enseignement, le point de départ étant la première année où débute l'enseignement formel de la lecture, de l'écriture et des mathématiques. Au moment de l'évaluation, l'âge moyen des élèves ne peut être inférieur à 9.5 ans. Aucune limite supérieure d'âge n'est fixée. À l'intérieur des groupes ainsi définis, des variations d'âge sont inévitables. Elles sont notamment

17. Pour plus de détails sur les modèles de la réponse à l'item, on se reportera à Lafontaine et Raïche (2011) ou à Bertrand et Blais (2004).

dues aux politiques de redoublement et au fait que dans certains systèmes, il faut avoir 6 ou 7 ans accomplis pour entrer en 1^{re} primaire.

Pour les systèmes éducatifs qui pratiquent peu le redoublement, le choix d'une approche à âge ou à niveau d'études constant change en réalité peu de choses, puisque les élèves d'un même niveau ont quasi le même âge. Pour les systèmes qui recourent fréquemment au redoublement, comme la Belgique, la France, le Luxembourg, l'Espagne, l'approche à âge constant peut faire une différence. Les élèves étant répartis sur plusieurs niveaux d'étude, la dispersion des résultats a tendance à être plus élevée dans les études à âge constant que dans les études à grade constant, et les indicateurs d'équité – ceux liés au sexe notamment – peuvent s'en trouver affectés, du fait que les garçons redoublent davantage que les filles. Par exemple, en mathématiques et en sciences en Communauté française de Belgique, les moyennes PISA des filles et des garçons sont quasi identiques, ce qui donnerait à penser, de prime abord, que les filles ont comblé leur retard historique dans les disciplines scientifiques. Une analyse plus approfondie (Baye et al., 2009) révèle qu'à l'intérieur de chacune des années d'études, et donc à parcours scolaire identique, les garçons ont des résultats systématiquement supérieurs à ceux des filles : l'équité est donc toute relative... et le message politique que l'on pourrait tirer de cet indicateur se révèle fortement orienté par le choix de la population de référence¹⁸.

Choix des textes et difficulté des tâches

Les références d'articles qui se sont émus des risques de biais culturels dans le choix des textes utilisés dans les enquêtes internationales ne manquent pas (Blum, Goldstein & Guérin-Pace, 2001 ; Bonnet, Deams & Glopper, 2003 ; Romainville, 2002). Ce n'est pas sur cette polémique que nous souhaitons revenir ici, mais sur un impact moins connu que peut avoir le choix des textes et du niveau de difficulté des tâches. Beaucoup de personnes ignorent que les textes supports de l'évaluation dans PIRLS, pour des élèves de 10 ans en moyenne, sont en moyenne deux fois plus longs que les textes de PISA (élèves de 15 ans). Ceci a de quoi surprendre. Alors que dans PISA, l'option de faire varier la longueur des textes et de limiter le nombre d'items posés par texte a été choisie, dans PIRLS un choix inverse a été effectué : tous les textes sont longs (au moins deux pages), les textes sont nettement moins nombreux et le nombre de questions posées par texte est élevé (jusqu'à 12 ou 13 items par texte). Ce choix n'est pas anodin¹⁹.

18. Lafontaine et Monseur (2009) ont montré que les indicateurs d'équité garçons-filles avaient tendance à être affectés par le choix d'une population d'âge dans les pays qui ne pratiquent pas la promotion automatique, en raison du fait que les garçons redoublent davantage que les filles.

19. Sur le plan de la mesure, l'option choisie dans PIRLS est difficile à comprendre dans la mesure où elle accroît le risque de dépendance entre items et diminue la part d'information

Nous faisons l'hypothèse que les élèves de 10 ans peuvent être rebutés par la longueur des textes (la longueur à cet âge représente une réelle difficulté), certains peuvent hésiter ou avoir du mal à lire tout le texte, et surtout risquent de répondre de mémoire aux questions plutôt que de relire le texte. Ceci risque d'affecter en particulier les élèves peu performants et les systèmes éducatifs qui comptent une proportion importante d'élèves peu performants en lecture. Si nous ne pouvons mettre cette hypothèse à l'épreuve à partir des données PIRLS, nous ne sommes toutefois pas totalement dépourvus pour approcher cette question ; les données de PISA 2009 apportent des éléments de réflexion utiles à cet égard.

Pour répondre à une préoccupation des systèmes éducatifs obtenant des performances largement en dessous de la moyenne des pays de l'OCDE, en particulier des pays dits partenaires, non-membres de l'OCDE participant à PISA, il a été décidé, dans PISA 2009, d'étendre l'échelle vers le bas et donc d'ajouter des unités très simples, constituées de textes courts et d'un nombre limité d'items (deux ou trois). Les pays qui lors de PISA 2000 avaient obtenu un score inférieur à 450 ont eu la possibilité d'adjoindre ce « carnet facile » à l'évaluation PISA 2009. On peut dès lors, pour ces pays où une proportion importante voire la majorité des jeunes se retrouvent dans les niveaux les plus élémentaires, distinguer ceux qui ont des compétences assez élémentaires ou vraiment rudimentaires en lecture.

Au-delà de l'intérêt que peut présenter cette approche plus fine sur le plan des résultats, on peut se demander quels sont les effets produits sur le niveau de performances des pays par un changement qui peut paraître mineur. Cette question peut être approchée en calculant par pays la corrélation entre son rendement moyen en 2000 et le progrès enregistré entre 2000 et 2009. Cette corrélation est robuste : -0.69 . Les pays qui avaient les scores les plus faibles en 2000 ont donc progressé bien davantage que les pays qui avaient les scores les plus élevés. Ceci peut évidemment tenir à de nombreux facteurs : la marge de progrès des pays les plus faibles est plus grande ; en outre, les systèmes qui avaient enregistré les performances les plus faibles en 2000 avaient davantage de raisons de s'inquiéter de la situation et se sont sans doute montrés plus actifs dans la recherche de solutions. Mais l'ampleur de la corrélation donne à penser que le fait d'inclure dans l'évaluation des unités « de base » vraiment plus faciles permet à des élèves de mettre en œuvre des compétences, fussent-elles rudimentaires, plutôt que de s'abstenir ou de répondre au hasard. Ce choix de placer plus ou moins haut le seuil d'entrée de l'évaluation, bien plus que l'origine culturelle ou linguistique des textes, nous semble de nature à jouer sur les différences de performances entre pays.

apportée par chaque item. En termes concrets, si un élève n'a pas compris un texte ou n'est pas du tout intéressé à lire un des textes, il risque d'échouer à un nombre important d'items. S'il n'y a que trois ou quatre items par texte, ce risque est plus limité.

LES SAVOIRS PRODUITS PAR LES ENQUÊTES INTERNATIONALES

Un éventail de savoirs, de formes variées, pour des usages multiples

Le corpus qui devrait être mobilisé pour approcher d'une manière empirique la question des savoirs produits par les enquêtes internationales et de leurs usages est, vu le nombre de pays impliqué, gigantesque. L'espace limité qui est à disposition ici pour discuter de ces questions ne permettra que d'ouvrir quelques pistes de réflexion. À la différence d'autres types d'évaluation, les évaluations internationales concernent potentiellement un large éventail d'« usagers » : élèves, enseignants, chefs d'établissement, décideurs en matière de politiques éducatives (ministres, parlementaires, responsables politiques et administratifs), syndicats, associations de parents, groupes militants, chercheurs. Si on prend l'exemple de PISA, la plus médiatisée des enquêtes internationales, il ne fait pas de doute que dans certains systèmes éducatifs, tels ceux de l'Allemagne, la Communauté française de Belgique, la Pologne, un large éventail d'acteurs se sont emparés, à un moment ou à un autre, des résultats de PISA pour des usages multiformes.

À cette diversité de publics-cibles potentiels correspond une diversité de canaux de diffusion. La comparaison sur internet des pages respectives de l'IEA²⁰ et de PISA²¹ est à cet égard instructive. Alors que l'IEA privilégie des documents de forme classique (typiquement le cadre d'évaluation, l'encyclopédie, le rapport international et le rapport technique), PISA développe, à côté des cadres d'évaluation, des rapports internationaux et techniques, une panoplie de documents plus courts (résumés, *executive summary*), ciblés sur une thématique précise (*PISA à la loupe*)²², censés répondre aux préoccupations plus politiques ou citoyennes, et des documents de recherche (*working papers*) qui intéresseront davantage un public expert. Des canaux de diffusion actuels tels que YouTube et Twitter sont également utilisés. Si les stratégies de l'OCDE sont plus affinées²³ et plus appuyées que celles de l'IEA, il n'en reste pas moins que les deux organisations internationales partagent des ambitions communes. Au travers des enquêtes internationales, il s'agit de toucher à la fois la communauté

20. <http://www.iea.nl> et <http://timssandpirls.bc.edu>

21. <http://www.pisa.oecd.org>

22. Documents de 4 pages qui font le tour d'une question. On citera à titre d'exemple : « Redoublement et transfert des élèves : quel impact pour les systèmes d'éducation ? », « Enseignement privé : quels avantages ? Et pour qui ? », « Les élèves d'aujourd'hui lisent-ils par plaisir ? ». Disponibles sur : http://www.oecd.org/document/24/0,3746,en_32252351_32235731_38378840_1_1_1_1,00.html

23. Les moyens de l'OCDE sont aussi sans commune mesure avec ceux, nettement plus modestes, de l'IEA.

scientifique, les citoyens et les décideurs politiques, autrement dit de produire des savoirs descriptifs ou explicatifs pour la recherche et des savoirs qui ambitionnent de guider l'action politique ou publique ; ces derniers auront alors une portée plus normative.

Des savoirs pour la recherche

Dans le domaine de l'évaluation, les enquêtes internationales, prototype du paradigme de la mesure, figurent parmi les types d'évaluation qui produisent des résultats de portée générale, transposables et généralisables. Au-delà des rapports directement édités par les organismes internationaux (OCDE, IEA), se développent de nombreux travaux de recherche à portée internationale ou plus locale, qui tirent parti de l'accès libre et gratuit aux bases de données internationales pour des analyses secondaires. L'activité de recherche à partir des bases de données internationales est prospère et attire aussi bien chercheurs en sciences de l'éducation qu'économistes ou sociologues de l'éducation. Comme le soulignent Felouzis et Hanhart (2011), « leur visée est évaluative, mais leurs usages peuvent être scientifiques » (p. 13).

Ainsi, les enquêtes internationales ont largement contribué à montrer que les systèmes « compréhensifs », peu ségrégués au début du secondaire (pas de filières, faible taux de redoublement) sont sensiblement plus équitables que ceux qui gèrent l'hétérogénéité des groupes d'élèves en les répartissant en groupes distincts (Monseur & Lafontaine, 2009). Dans un tout autre domaine, une recherche quasi fondamentale est venue se greffer sur PISA 2009²⁴ : la lecture en ligne fait-elle appel à d'autres habiletés que la lecture sur support papier ? Quel rôle joue la navigation ? Le rapport consacré aux résultats de la lecture en ligne met en évidence que les performances en lecture électronique s'expliquent bien entendu largement par les capacités de lecture générales, mais pas seulement. La composante navigation joue aussi un rôle essentiel (OCDE, 2011a). Le savoir produit dans le présent cas s'appuie sur un échantillon sans commune mesure avec celui d'une recherche expérimentale classique, et se révèle non seulement utile sur le plan de la recherche fondamentale, mais aussi sur le plan didactique.

Une double polarité : entre constat et norme

Comme l'indiquent Felouzis et Hanhart (2011), « les ambiguïtés de l'évaluation entre pratique scientifique et usages normatifs peuvent être vues non pas comme un problème à résoudre, mais comme une de ses dimensions constitutives » (p. 13). Les savoirs issus des enquêtes internationales oscillent

24. Dans PISA 2009, à côté du test classique papier-crayon, les pays, à titre d'option, pouvaient choisir de tester les capacités de lecture en ligne de leurs élèves de 15 ans.

en permanence entre constat et norme. De prime abord, il s'agit de classer les systèmes éducatifs, de comparer leurs performances, d'appréhender leur équité, de tenter ensuite d'expliquer tant l'efficacité que l'équité à la lumière de nombreux facteurs contextuels. Explicitement, en surface, la tonalité du discours est descriptive ou explicative. Mais la norme ou la tentation de glisser du constat à la norme n'est jamais loin. À titre d'illustration, voici comment se conclut le *PISA à la loupe* consacré au redoublement et au transfert des élèves :

Pour conclure : certaines politiques visant à regrouper les élèves selon leur potentiel académique, leurs intérêts ou leur comportement, telles que le redoublement ou le transfert des élèves vers d'autres établissements, peuvent être coûteuses pour les systèmes d'éducation, tout en n'étant généralement pas associées à une amélioration de la performance des élèves ou à un renforcement de l'égalité des chances dans l'apprentissage. (OCDE, 2011b, p. 4)

Sous le constat – dont il ne fait pas de doute qu'il s'appuie sur de solides évidences empiriques – on sent poindre l'incitation à remettre en cause les politiques de redoublement.

Des savoirs pour quels usages ?

Les enquêtes internationales ont l'ambition – c'est un truisme de l'affirmer – de produire des savoirs utiles pour l'action et le pilotage des systèmes éducatifs. Les décideurs politiques (*stakeholders*) sont leur cible privilégiée. À cet égard, elles s'inscrivent dans une perspective rationaliste, planificatrice – voire technocrate – qui veut fonder la régulation des systèmes éducatifs sur des indicateurs de performances et d'équité mesurables.

Les enquêtes internationales s'inscrivent ainsi dans le mouvement plus large qui tend à asseoir la définition et la régulation des politiques publiques sur des indicateurs et des évaluations et à lier savoirs et action publique. Les enquêtes internationales présentent cependant des spécificités qui les distinguent d'autres formes d'évaluation. Comme le développe Steiner-Khamsi (2003),

les classements et palmarès des études types OCDE et IEA constituent une forme mesurable, aisément accessible [...] de forme de « rationalité scientifique » qui permet aux décideurs politiques dans le domaine de l'éducation de convaincre le grand public lorsqu'ils planifient ou suspendent une réforme d'une certaine ampleur. (p. 2)

Mais il y a davantage. « L'externalisation, ou la référence à d'autres systèmes éducatifs, fonctionne comme une source d'autorité » (p. 2). Les enquêtes internationales sont de plus en plus utilisées pour juger de l'efficacité et de l'équité des systèmes éducatifs, avec l'idée qu'il est possible d'emprunter ou de s'inspirer de l'exemple des systèmes éducatifs les plus performants et les plus équitables. Selon Cattonar *et al.* (2009) :

Dans l'espace public, le premier et principal effet du programme PISA est d'ordre cognitif : il transforme le cadre cognitif dans lequel on pense les questions d'éducation en les inscrivant dans un nouvel espace dont l'échelle est internationale et dont les dimensions sont fondées sur des mesures objectivées de résultats. (p. 151)

Les savoirs issus des enquêtes comparatives contribuent ainsi à légitimer des actions et des politiques éducatives. PISA est très vite devenu, sur la scène internationale, une référence qui a pu être convoquée soit pour légitimer des réformes engagées antérieurement, soit pour légitimer des actions en sens diamétralement opposés. Ainsi, en Communauté française de Belgique, PISA a été utilisé aussi bien par un groupe militant comme l'Association pour une école démocratique (Aped)²⁵ pour appuyer ses revendications telles que supprimer les filières et instaurer une carte scolaire (Hirtt, 2010), que par le Mouvement réformateur (libéral) pour justifier un renforcement de l'autonomie des écoles.

Au cœur de ce processus de légitimation s'opère un rapprochement des savoirs et de l'action, qui peut aller jusqu'à un brouillage complet des frontières²⁶. Les savoirs tombés ainsi dans le domaine public ont tendance à être réifiés, présentés comme neutres ou objectifs ; les données sont souvent surinterprétées ou isolées de leur contexte de production ou d'interprétation (le système d'où elles proviennent). PISA peut ainsi devenir, dans les mains de certains, la mesure de toute chose, alors qu'elle n'est qu'une mesure parmi d'autres.

Si le « déferlement d'une rhétorique considérant l'évaluation comme l'outil cardinal de la bonne gouvernance » (Mons & Crahay, 2011, p. 77) ne fait guère de doute, tout autre est la question, au-delà des discours et des cadres de pensée, des effets réels de ces évaluations sur les politiques éducatives. Selon les mêmes auteurs, il existe un « écart manifeste, en ce qui concerne la formulation des politiques, entre la rhétorique de la responsabilisation et la réalité des dispositifs mis en œuvre » (p. 91).

EN GUISE DE CONCLUSION : LA DÉLICATE POSTURE DU CHERCHEUR EN ÉVALUATION

Comme nous l'avons montré dans la première partie de cette contribution, les enquêtes internationales portant sur les acquis des élèves ont des caractéristiques et des limites interprétatives qui tiennent à la fois aux théories

25. <http://www.skolo.org/>

26. Un exemple un peu anecdotique illustre bien ce brouillage des frontières entre savoirs et politique. Sur un plateau de télévision consacré aux résultats de PISA (RTBF, La Une, *Mise au point*, 11 décembre 2011), les différents hommes politiques s'étaient mués en chercheurs... commentant abondamment l'enquête et ses résultats, tandis que le journaliste multipliait les questions au chercheur sur les réformes et actions à entreprendre.

du domaine évalué, au paradigme dont elles relèvent et aux modèles de mesure mobilisés, ainsi qu'à la méthodologie mise en œuvre. Les indicateurs ne sont ni des évidences ni des reflets de la réalité ; ce sont des construits. Les limites et la prudence avec laquelle il faut prendre certains indicateurs sont souvent balayées lorsque les résultats servent d'instruments pour un usage normatif – ce à quoi les évaluations ne peuvent échapper. Dans la seconde partie de la contribution, nous nous sommes fait brièvement l'écho d'approches plus sociologiques qui analysent le lien entre les évaluations internationales et les politiques publiques.

Contrairement au sociologue qui prend les évaluations pour objet d'analyse (voir par exemple Martuccelli, 2010), les chercheurs en évaluation que nous sommes, sauf à changer de métier, ne peuvent se soustraire à la question du « comment faire de l'évaluation comparée des acquis ? » Si les évaluations internationales jouent aujourd'hui ce rôle de matrice cognitive et normative jugé hégémonique par certains (Bulle, 2010 ; Maroy & Mangez, 2011 ; Martuccelli, 2010), quelle posture épistémologique peuvent adopter les chercheurs en évaluation en charge de telles enquêtes, qui préviennent ou limitent les risques de voir ces enquêtes fonctionner comme une doxa ou de fausses évidences ? Ceux-ci doivent d'abord, à notre sens, veiller avec fermeté aux limites interprétatives des résultats de l'enquête. Si l'on veut évaluer dans quelle mesure un système éducatif atteint ses propres objectifs tels que définis dans les instructions officielles, les enquêtes internationales ne sont pas l'instrument adéquat ; il faut le dire sans détour. Le chercheur n'échappera pas à l'exercice périlleux de communication des résultats des enquêtes internationales vers des publics de non-spécialistes ; ceci suppose une simplification qui doit néanmoins s'opérer sans trahir un certain nombre de principes. Avec la subjectivité assumée de notre expérience d'acteur averti, nous voudrions attirer l'attention sur les risques qui nous paraissent peser sur l'exercice, ou sur les tensions que nous avons ressenties lorsque nous avons été amenés à le pratiquer. Un risque, majeur lorsqu'on interagit avec la presse, est que le projecteur soit placé sur le seul classement, voire sur le rang précis occupé par le pays dans le classement. Le chercheur se doit de rappeler que les écarts entre pays sont somme toute assez minimes si on les compare aux écarts entre élèves à l'intérieur des pays. Le chercheur aura aussi à naviguer entre ceux qui, face à un piètre ou un trop bon classement²⁷, seront tentés de « casser le thermomètre », et, à l'opposé, ceux qui feraient volontiers de PISA la mesure de toute chose ou encore rêveraient de recettes ou bonnes pratiques universellement applicables.

La demande+ est souvent forte pour pousser le chercheur à dépasser les constats et à trancher des questions de politique éducative sur la base de son

27. Le Québec par exemple, très surpris par son bon classement dans les enquêtes PISA et PIRLS, est porté à contester la nature des enquêtes internationales. De bons résultats vont évidemment à l'encontre des intérêts de ceux qui souhaitent réformer le système ou les pratiques d'enseignement.

expertise des enquêtes internationales. Exemple : « Si les pays sans filières et sans redoublement sont plus équitables (constat), alors faut-il supprimer les filières et le redoublement dans notre système éducatif (norme) ? » Face à cette croyance dans les vertus et les pouvoirs de la rationalité scientifique, le chercheur en évaluation devrait agir comme un garde-fou et s'en tenir à une des règles d'or de la recherche scientifique en sciences humaines. Pour reprendre la célèbre distinction de Max Weber (1959²⁸), la recherche empirique ne peut à elle seule fournir des réponses face à des questions qui relèvent autant de l'éthique ou de l'axiologie (valeurs, choix politique) que de la recherche scientifique (faits, constats). Il appartient au chercheur en évaluation qui est impliqué dans les enquêtes internationales d'être particulièrement vigilant sur ce point ; faire l'économie d'une réflexion sur les valeurs éducatives, ancrée dans les réalités sociopolitiques et culturelles, est en effet un écueil majeur dans le cadre globalisant des enquêtes internationales.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Baye, A., Fagnant, A., Hindryckx, G., Lafontaine, D., Matoul, A. & Quittre, V. (2009). *Les compétences des jeunes de 15 ans en Communauté française en sciences, en mathématiques et en lecture. Résultats de l'enquête PISA 2006* (Cahiers des sciences de l'éducation, 29-30). Liège : Université de Liège.
- Bertrand, R. & Blais, J.-G. (2004). *Modèles de mesure. L'apport de la théorie des réponses aux items*. Sainte-Foy, Québec : Presses de l'Université du Québec.
- Blum, A., Goldstein, H. & Guérin-Pace, F. (2001). International Adult Literacy Survey (IALS) : An analysis of international comparisons of adult literacy. *Assessment in Education*, 8(2), 225-246.
- Bonnet, G., Deams, F. & Glopper, C. (2003). *Culturally Balanced Assessment of Reading (C-Bar)*. Paris : Ministère de la jeunesse, de l'éducation nationale et de la recherche, Direction de l'évaluation et de la prospective.
- Bulle, N. (2010). *L'imaginaire réformateur – PISA et les politiques de l'école*. Consulté le 5 août 2011 dans <http://skhole.fr/l-imaginaire-r%C3%A9formateur-pisa-et-les-politiques-de-l-%C3%A9cole>
- Cattonar, B., Mangez, E., Delvaux, B., Mangez, C. & Maroy, C. (2009). *Usage et circulation des enquêtes PISA en Communauté française de Belgique* (Rapport de recherche). KnowandPol, Belgian Education Team, R&D EU Sixth Framework Program. Consulté le 6 décembre 2011 dans <http://www.knowandpol.eu>
- De Ketele, J.-M. (2010). Faces visibles et faces cachées des classements internationaux : une tentative de modélisation des tensions dans l'enseignement obligatoire. *Revue internationale d'éducation*, 54, 39-49.

28. Traduction de deux conférences originales en allemand de 1917 et 1919.

- De Ketele, J.-M. & Gérard, F. M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.
- Demonty, I. & Fagnant, A. (2004). *Évaluation de la culture mathématique des jeunes de 15 ans : document à l'attention des professeurs de mathématiques des 1^{er} et 2^e degrés de l'enseignement secondaire*. Bruxelles : Ministère de la Communauté Française, Service Général du Pilotage du Système Éducatif.
- Felouzis, G. & Hanhart, S. (2011). Politiques éducatives et évaluation : nouvelles tendances, nouveaux acteurs. In G. Felouzis & S. Hanhart (Éd.), *Gouverner l'éducation par les nombres ? Usages, débats et controverses* (pp. 7-31). Bruxelles : De Boeck.
- Grek, S. (2009). Governing by numbers : the PISA 'effect' in Europe. *Journal of Education Policy*, 24(1), 23-37.
- Hirtt, N. (2010). *PISA ? C'est toujours la catastrophe*. Consulté le 4 août 2011 dans <http://www.skolo.org/spip.php?article1276>
- Kennedy, A. M., Mullis, I. V. S., Martin, M. O. & Trong, K. L. (2007). *PIRLS 2006 Encyclopedia. A Guide to Reading Education in the Forty PIRLS 2006 Countries*. Chestnut Hill, Massachusetts : TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Lafontaine, D. (2010). Les standards internationaux en éducation : la place des situations et des contextes. In D. Masciotra, F. Medzo & P. Jonnaert (Éd.), *Vers une approche située en éducation. Réflexions, pratiques, recherches et standards* (Cahiers scientifiques de l'ACFAS, 111, pp. 159-174). Montréal : Association francophone pour le savoir.
- Lafontaine, D. & Monseur, C. (2009). Gender Gap in Comparative Studies of Reading Comprehension : to what extent do the test characteristics make a difference ? *European Educational Research Journal*, 8(1), 69-79.
- Lafontaine, D. & Raïche, G. (2011). Principes méthodologiques et techniques des enquêtes internationales. *Mesure et évaluation en éducation*, 34(2), 27-54.
- Maroy, C. & Mangez, C. (2011). La construction des politiques d'évaluation et de pilotage du système scolaire en Belgique francophone : nouveau paradigme politique et médiation des experts. In G. Felouzis & S. Hanhart (Éd.), *Gouverner l'éducation par les nombres ? Usages, débats et controverses* (pp. 53-75). Bruxelles : De Boeck.
- Martin, M. O., Mullis, I. V. S. & Kennedy, A. M. (2007). *PIRLS 2006 Technical Report*. Chestnut Hill, Massachusetts : TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Martuccelli, D. (2010). Critique de la philosophie de l'évaluation. *Cahiers internationaux de sociologie*, 128-129, 27-52.
- Mons, N. & Crahay, M. (2011). L'évaluation des performances scolaires des élèves : un instrument d'évaluation des politiques éducatives ? In

- G. Felouzis & S. Hanhart (Éd.), *Gouverner l'éducation par les nombres ? Usages, débats et controverses* (pp. 77-98). Bruxelles : De Boeck.
- Monseur, C. & Lafontaine, D. (2009). L'organisation des systèmes éducatifs : quel impact sur l'efficacité et l'équité ? In V. Dupriez & X. Dumay (Éd.), *L'efficacité en éducation, promesses et zones d'ombre* (pp. 141-165). Bruxelles : De Boeck.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L. & Sainsbury, M. (2009). *PIRLS 2011 Assessment framework*. Chestnut Hill, Massachusetts : TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Erberber, E. (2005). *TIMSS 2007 Assessment framework*. Chestnut Hill, Massachusetts : TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Organisation de coopération et de développement économiques. (1999). *Mesurer les compétences et les connaissances des élèves : un nouveau cadre d'évaluation*. Paris : Organisation de coopération et de développement économiques.
- Organisation de coopération et de développement économiques. (2009a). *PISA 2006 Technical Report*. Paris : Organisation de coopération et de développement économiques.
- Organisation de coopération et de développement économiques. (2009b). *Le cadre d'évaluation de PISA 2009 : les compétences en compréhension de l'écrit, en mathématiques et sciences*. Paris : Organisation de coopération et de développement économiques.
- Organisation de coopération et de développement économiques. (2011a). *PISA 2009 Results : Students On Line : Digital Technologies and Performance* (Vol. 6). Paris : Organisation de coopération et de développement économiques.
- Organisation de coopération et de développement économiques. (2011b). Redoublement et transfert des élèves : quel impact pour les systèmes d'éducation. *PISA à la loupe*, 6, 1-4.
- Robitaille, D. F., Schmidt, W. H., Raizen, S., Mc Knight, C., Britton, E. & Nicol, C. (Éd.). (1993). *Curriculum frame-works for mathematics and science*. Vancouver : Pacific Educational Press.
- Romainville, M. (2002). Du bon usage de PISA. *La Revue Nouvelle*, 3-4, 86-99.
- Rosier, M. J. & Keeves, J. P. (Éd.). (1991). *The IEA Study of Science I : Science Education and Curricula in Twenty-Three Countries*. Oxford : Pergamon Press.
- Steiner-Khamsi, G. (2003). The politics of league-tables. *Journal of Social Science Education*, 1-2003. Consulté le 6 décembre 2011 dans <http://www.jsse.org/2003/2003-1/tables-khamsi.htm>
- Travers, K. J. & Westbury, I. (1989). *The IEA Study of Mathematics I : Analysis of Mathematics Curricula*. Oxford : Pergamon Press.
- Weber, M. (1959). *Le Savant et le Politique*. Paris : Plon.

Les évaluateurs, des épistémologues en actes ?

**Claire Tourmen & Patrick Mayen
AgroSup Dijon**

Dans le cadre de la problématique de cet ouvrage, nous avons décidé d'interroger l'épistémologie de l'évaluation à travers celle(s) des évaluateurs. Si l'on prend « épistémologie » dans son sens étymologique, il s'agit de la « science du connaître », signification qui s'est ensuite dédoublée en « théorie de la connaissance » et en « discours sur la science » (Barreau, 1992). Partant du point de vue qui est le nôtre, l'analyse du travail d'évaluateurs en situations réelles, nous nous sommes demandé quel était le rapport des évaluateurs aux phénomènes sur lesquels ils ont à enquêter et qu'ils ont à juger, s'ils disposaient d'une quelconque « théorie de la connaissance » à leur égard, voire s'ils discouaient sur elle et, surtout, dans une perspective pragmatique, s'ils en faisaient quelque chose dans l'action et pour l'action. C'était là une des hypothèses formulées par Shadish, Cook et Leviton (1991) dans leur livre fondateur sur les théories de l'évaluation : les évaluateurs mobiliseraient, entre autres, une théorie de « la construction du savoir » et de « la manière dont on apprend des programmes » : « Comment est-ce que je connais tout ceci ? Quel type de réponse est digne de confiance ? Qu'est-ce qui inspire de la confiance ? » (p. 35).

Ceci nous a amenés à nous questionner sur le rôle et la place de l'expérience dans les pratiques d'évaluation. En effet, les évaluateurs font l'expérience de la situation (phénomènes à évaluer et conditions pour le faire), soit l'expérience *processus*¹, à savoir un enchaînement d'événements vécus dans leur succession (Mayen, 2009). Les évaluateurs ont aussi l'expérience passée des situations

1. « L'expérience comme événement correspond à ce qu'on peut appeler "faire l'expérience" de quelque chose, d'une situation, autrement dit de la transaction d'une personne avec une partie de son environnement. » (Mayen, 2009, p. 767).

d'évaluation et des phénomènes à évaluer, soit l'expérience *produit* au sens de « composition » de « manières d'être, de penser et de faire » qui deviennent autant de « prédispositions » ou « potentialités » pour vivre des expériences ultérieures (Mayen, 2009, p. 770). Dans la démarche scientifique, l'expérience *processus* comme l'expérience *produit* sont suspectes parce qu'elles relèvent d'une trop grande proximité avec les phénomènes, proximité avec laquelle il est demandé de rompre (Bachelard, 1971). Si la démarche scientifique moderne s'est construite sur une mise à distance contrôlée de l'expérience des phénomènes, en est-il de même dans les pratiques professionnelles d'évaluation ? Peut-on identifier la présence d'actes et d'activités épistémologiques ? Vergnaud (1992) propose de définir l'épistémologie « dans un sens volontairement restreint du terme, [comme] la relation qu'entretiennent un savoir et un savoir-faire avec les problèmes pratiques et théoriques auxquels il apporte une réponse » (p. 23). Cette définition, même si elle prend la précaution de souligner son caractère restrictif, intéresse cependant les cas qui nous occupent : en cours d'action, l'épistémologie a-t-elle sa place pour contribuer à l'efficacité de l'action ?

Notre réflexion s'est principalement appuyée sur deux recherches concernant des activités d'évaluation dans le secteur éducatif : des jurys de validation des acquis de l'expérience (VAE) et des épreuves d'évaluation en situation professionnelle dans l'enseignement secondaire agricole. Nous ferons aussi référence à d'autres études sur l'activité d'évaluateurs que nous-mêmes ou d'autres ont pu conduire.

Nos observations montrent que les évaluateurs que nous avons rencontrés : 1) ne se coupent pas de l'expérience immédiate et passée des phénomènes évalués, mais qu'ils s'y enracinent et l'utilisent comme ressource ; 2) ont aussi appris à s'en méfier et à l'encadrer, guidés par ce que nous avons convenu d'appeler différentes « épistémologies en actes ».

LES DÉMARCHES PROFESSIONNELLES D'ÉVALUATION, UNE TENTATIVE DE RUPTURE AVEC LES JUGEMENTS SPONTANÉS DE L'EXPÉRIENCE

La littérature en évaluation s'est construite, depuis les années 1950, sur le projet de fonder et d'outiller une pratique professionnelle de l'évaluation en rupture avec les pratiques spontanées de jugement de l'expérience quotidienne. Cette littérature peut se lire comme une tentative de se prémunir contre « l'arbitraire des jugements » (Figari, 1994) non justifiés, potentiellement biaisés et proférés avec autorité dans le cours de l'action.

Énoncer des jugements, un acte quotidien

Évaluer est un acte courant et spontané que chacun d'entre nous réalise au cours de l'expérience *processus*, acte qui consiste à juger, estimer la

valeur de quelqu'un ou quelque chose (Barbier, 1985). Dans sa forme la plus réduite et la plus courante, évaluer se réalise par un acte de langage qui consiste à qualifier – attribuer une qualité à – un événement du monde, à en apprécier le poids, la valeur, à lui attribuer estime, admiration ou au contraire mépris (Dewey, 1993). Dans ce cas, « le jugement de valeur ne s'explique qu'à travers son énoncé, sa formulation » (Barbier, 1985, p. 33). Quand la situation est déterminée et non problématique, les jugements vont de soi. À l'inverse, quand les situations sont problématiques et indéterminées, elles engagent les individus dans des démarches « d'enquête » (Dewey, 1993). Dewey souligne la fréquence de ces « enquêtes et jugements de sens commun » et autres « délibérations de la vie quotidienne » où nos décisions sont pour partie fondées sur des enquêtes – et de la connaissance – et non uniquement sur du « caprice » ou de « l'habitude aveugle » (1993).

Énoncer des jugements, un acte contestable

Pourtant, la légitimité de ces « enquêtes et jugements de sens commun » peut être contestée. En effet, si juger est un acte qui a des conséquences pratiques, qui a autorité et compétence pour juger ? Celui qui juge n'est-il jamais victime de ses « habitudes » ou « impulsions » ? Les auteurs en évaluation se méfient des jugements de pratique quotidiens et spontanés issus d'enquêtes souvent rapides. Selon Nadeau, « ces évaluations journalières sont informelles, ce qui ne veut pas dire qu'elles soient inexactes, en ce sens qu'elles sont le plus souvent basées sur des impressions vagues et diffuses et sur des normes intuitives » (1988, p. 35). Les jugements spontanés courent plusieurs risques : n'être basés que sur des informations limitées – voire erronées – sur les phénomènes évalués, voire sur la seule intuition de l'évaluateur ; n'en estimer la valeur qu'au regard de référents et normes implicites voire ignorés, ou de référents limités au seul point de vue et au seul intérêt d'une des parties. Les jugements spontanés sont en effet empreints d'affects (biais d'empathie ou d'antipathie), d'intérêts particuliers et de tentatives de détournement de la démarche d'évaluation à leur profit (Lemieux, 2002), de l'influence de normes non conscientes et pourtant agissantes (telle la « norme d'internalité » mise à jour par des travaux en psychologie sociale [Bressoux & Pansu, 2003]), d'une connaissance limitée des phénomènes, de phénomènes de distorsion du jugement par « effet d'attente » ou « effet d'ancre » (Abernot, 1988), etc.

Énoncer des jugements, un acte professionnel ?

Pour lutter contre la tendance à juger trop vite, des méthodes ont été construites. Ainsi sont nées des démarches professionnelles d'évaluation, appelant la mise en place d'enquêtes rigoureuses, inspirées (et se revendiquant parfois) des démarches d'enquêtes scientifiques. De nombreux auteurs estiment qu'une démarche d'évaluation professionnelle se doit

d'être transparente, négociée, rigoureuse. L'évaluation de programme s'est constituée, à partir des années 1960 en Amérique du Nord, en empruntant des méthodes d'enquête aux sciences sociales et même aux sciences de la nature (« expérimentation sociale »), de nombreux évaluateurs étant eux-mêmes chercheurs et certains se revendiquant d'une « recherche évaluative ». De leur côté, des chercheurs en sciences de l'éducation ont tenté de construire des protocoles pour mesurer scientifiquement les apprentissages, comme en éduométrie, par un travail sur les conditions de fidélité et de généralisation de la mesure. La construction des épreuves d'évaluation et leur contrôle selon certaines techniques permettraient ainsi, selon Cardinet (2003), d'assurer la « scientificité » des mesures effectuées. D'où notre questionnement : les professionnels de l'évaluation sont-ils conscients de tout ceci lorsqu'ils évaluent ? Mènent-ils des actes épistémologiques ? En parlent-ils ?

L'ANALYSE DU TRAVAIL D'ÉVALUATEURS DANS PLUSIEURS CONTEXTES

À la lumière de ce questionnement, nous avons repris des données issues de plusieurs recherches conduites dans une perspective d'analyse du travail en didactique professionnelle. Nous avons en effet observé le travail d'évaluateurs en situations réelles. Il s'agit tout d'abord de membres de jurys de VAE, pour des diplômés de différents niveaux, 10 jurys pour le secondaire et 13 jurys pour le supérieur (pour plus de détails sur les terrains, voir Mayen, Mayeux & Savoyant, 2006 ; Mayen & Tourmen, 2009). Nous avons observé et analysé le travail des jurys dans son intégralité, depuis les premiers moments de l'examen d'un dossier jusqu'à la délibération finale, en passant par les entretiens avec les candidats.

Nous avons ensuite observé un jury lors d'une épreuve pratique (en zootechnie) dans l'enseignement professionnel agricole (Tourmen & Deblay, 2011). Il s'agissait d'une épreuve de contrôle continu en seconde professionnelle « Production Animale », réalisée dans l'exploitation d'un lycée agricole. Dix élèves se sont succédé pour effectuer des tâches tirées au sort sous le regard et les questions de deux évaluateurs, une enseignante du lycée et un professionnel de l'élevage bovin. Ceux-ci ont ensuite délibéré pour noter les élèves au moyen d'une grille de critères.

Dans les deux cas, les méthodes de recueil de données étaient inspirées des méthodes d'analyse du travail utilisées en ergonomie (Leplat, 1997) : analyse de la tâche, observations de l'activité des évaluateurs et des évalués, entretiens pré et post activité avec les différents acteurs. Nous avons, à chaque fois, enregistré et intégralement retranscrit les interactions des protagonistes, afin de pouvoir mener une analyse détaillée du déroulement de leur activité. Celle-ci a été analysée dans son déroulement temporel,

en prenant en compte d'une part la progression de l'évaluation (jugements prononcés, critères et indicateurs énoncés, à quel moment, progression temporelle du jugement, types d'arguments exposés et retenus ou rejetés, nature des référents mobilisés) et, d'autre part, l'activité des protagonistes (actions réalisées, buts énoncés et pris en compte, indicateurs de situation prélevés, raisonnements menés en action, invariants opératoires mobilisés dans l'action ainsi que ceux qui ont été énoncés). Nous avons été particulièrement attentifs aux « concepts en actes » et « théorèmes en actes » (ou « propositions tenues pour vraies sur le réel » [Pastré, Mayen & Vergnaud, 2006]) qui se manifestent de façon explicite ou implicite dans l'action évaluative. Les analyses ont ainsi été construites sur deux modèles de l'activité : une théorie de l'activité évaluative comme activité de production de jugements, qui consiste à estimer le degré de réalisation de certains critères à l'aide d'indicateurs et au regard de référents (Barbier, 1985 ; Figari, 1994) ; et une théorie de l'activité de travail comme ce qui est fait (gestes et paroles) en lien avec ce qui est pensé, perçu et projeté (intentions, raisonnements tenus en situation, invariants opératoires mobilisés).

RÉSULTATS : LES ÉVALUATEURS UTILISENT L'EXPÉRIENCE DES PHÉNOMÈNES COMME RESSOURCE

La prégnance des premiers jugements

On peut constater, sur tous les terrains, l'existence de jugements précoces et globaux sur les candidats évalués, ce qui pourrait relever de l'idée de jugements hâtifs et spontanés de la vie courante. Les deux évaluateurs observés lors de l'épreuve en lycée agricole commencent systématiquement les délibérations sur chaque élève par l'expression d'un jugement global sur sa prestation (« bien », « dans l'ensemble ça va »), de même que les membres de jurys de VAE commencent leurs délibérations en exprimant un premier jugement global sur le niveau d'un candidat. Il s'agit de la première série d'échanges entre les membres du jury. Le même phénomène a été observé chez des évaluateurs dans d'autres contextes, lors de l'évaluation de programmes et politiques publics (Tourmen, 2009).

D'autres études rapportent le cas d'évaluateurs qui semblent se fier de manière privilégiée à ces premiers jugements. Ainsi, Merle (2007) rapporte les propos de deux enseignants sur les oraux du baccalauréat : selon l'un, « on les laisse parler pendant 10 minutes s'ils ont quelque chose à dire. On voit assez vite les limites », et selon l'autre : « en 5 minutes, tout est dit [...] il est rare que l'impression de départ soit infirmée » (p. 138). Merle qualifie ces professeurs de « professeurs intégristes », à la différence d'autres professeurs, qu'il qualifie de « dubitatifs », car ils expriment davantage de doutes sur leurs propres jugements. Un phénomène similaire a été observé par Martin (2011)

chez des recruteurs en Bâtiment et travaux publics (dont l'activité a été analysée comme une évaluation des candidats) : un recruteur se fait sa première impression « quand une personne passe la porte », « avant de lui parler j'ai déjà mon idée », « la première impression est essentielle, elle ne trompe pas » (p. 33).

La mise à distance des premiers jugements

Les évaluateurs que nous avons observés ne s'arrêtent pas à ces premiers jugements ; ils les discutent, les mettent en doute, cherchent et échangent des arguments pour les renforcer, les nuancer, les réfuter. On peut considérer ces premiers énoncés comme des propositions mises en commun afin d'être discutées. À partir de ce qui apparaît comme hypothèse de jugement, le travail commence. On ne trouve, en VAE, aucun membre de jury qui aurait présenté un premier jugement définitif. Un ensemble d'indices en atteste : le nombre et la variété des modalisations, les intonations, les énoncés explicites qui donnent à ces énoncés un statut d'impression, de proposition de discussion... Ainsi, après l'énonciation des premiers jugements, les membres de jurys de VAE expriment leur accord ou leur désaccord (« je suis tout à fait d'accord sur la partie économique, sur ce que vous venez de dire »), rapportent leurs impressions à des indicateurs ou des référents et normes qui les fondent, nuancent leurs propos en les ramenant à leur point de vue et en admettant, implicitement, l'existence d'autres points de vue (« j'ai l'impression... », « j'ai trouvé que... »). Les évaluateurs en lycée professionnel soumettent eux aussi ces premiers jugements à un examen, en les rapportant à des observations qu'ils ont faites sur le candidat (« [elle] a su trouver la dose relativement facilement », « elle a passé un quart d'heure pour calculer la dose... ») et à des référents extérieurs, par exemple ce que les élèves sont censés savoir (« les maladies, ils l'ont pas encore vu en fait, tout ce qui est santé, en cours »), ou ce que d'autres élèves ont accompli. La prise en compte de ces éléments leur permet de valider, d'affiner et de nuancer les premiers jugements. Dans les cas observés², les premiers jugements n'étaient donc pas pris pour argent comptant mais mis à l'examen, affinés, nuancés, acquérant alors le statut d'hypothèse initiale de travail dans l'ensemble du processus d'évaluation.

Le recours à l'expérience antérieure des phénomènes

Dans tout le processus de jugement et de débat sur les jugements, l'activité des évaluateurs s'enracine dans l'expérience *produit* (à savoir leur expérience

2. Ce qui signifie que cela pourrait être différent dans d'autres contextes, ou chez d'autres évaluateurs.

passée des phénomènes évalués), ce qui lui confère une certaine validité. En effet, les évaluateurs ne sont pas étrangers au domaine évalué. Quand bien même ils ne connaîtraient pas un candidat, ils ont l'expérience d'autres candidats, du programme de formation, du métier et, dans la plupart des cas, ils ont l'expérience de l'évaluation. Les évaluateurs sont souvent choisis en raison de leur expertise, comme dans le cas de la VAE où les membres du jury sont sélectionnés en fonction d'une expertise académique et/ou professionnelle.

L'analyse montre ainsi que les évaluateurs en lycée agricole mobilisent des connaissances dans différents domaines :

- Les gestes et raisonnements attendus d'un futur professionnel (« elle a pas la réflexion d'un futur éleveur et de raisonner au niveau pratique et organisation du travail ») ;
- Le degré de difficulté des situations de travail, et celui que les élèves sont censés maîtriser à ce stade de leur formation (les vaches étaient dites « faciles » lors de l'épreuve) ;
- Les connaissances censées être acquises (« normalement elle le sait ») ;
- Les écarts aux normes professionnelles admissibles dans le métier (« on le fait tous, on le fait tous mais effectivement c'est... »).

C'est la mobilisation de toutes ces connaissances qui permet de donner sens à l'activité d'un élève donné, de lui attribuer une certaine valeur et de la pondérer au regard de différents critères (la réussite étant, par exemple, pondérée au regard de la facilité de la situation d'évaluation).

Les jurys de VAE font également référence à diverses connaissances, issues à la fois des référentiels et de leur propre expérience. Qu'il s'agisse d'interpréter les référentiels ou les productions des candidats, les membres des jurys se réfèrent, dans leurs échanges, à plusieurs univers d'expérience relevant du monde du travail et de la formation (Mayen, Mayeux & Savoyant, 2006). Bonnes pratiques, pratiques « acceptables », exigences des secteurs d'activité, relations entre engagement dans le travail et capacités techniques, etc., les connaissances mobilisées dans les discussions ne sont pas des connaissances générales ni, pour la plupart, académiques, mais sont le plus souvent nourries de rencontres, d'expériences de travail, de visites de stage... Ces connaissances sont, ainsi, pour partie issues des référentiels (métiers, compétences, formation) et pour partie de l'expérience des membres du jury. Les référentiels n'apparaissent que comme des propositions ouvertes. Les jurys doivent mobiliser des références externes aux référentiels pour, en quelque sorte, « référentialiser » les référentiels et leur donner du sens pour évaluer. De plus, les évaluateurs se réfèrent et enquêtent aussi sur l'expérience d'autres personnes en lien avec ces phénomènes. Allal et Mottier Lopez (2009) rapportent le cas d'enseignants qui enquêtent sur la perception que d'autres collègues ou même les parents ont de certains élèves, afin de trianguler leurs sources d'information et d'affermir leurs jugements, particulièrement dans le cas d'élèves dont le niveau est difficile à déterminer.

Les connaissances issues de l'expérience : des référents dans l'évaluation

Les connaissances issues de l'expérience des phénomènes évalués acquièrent un nouveau statut dans l'acte d'évaluation : elles deviennent des « référents », ce que Figari (2006) définit comme un ensemble d'éléments normatifs et informatifs permettant de qualifier l'état attendu d'un objet. Elles servent d'éléments de comparaison pour attribuer un sens et une valeur aux nouveaux phénomènes observés, un cas étant jugé à l'aune de la connaissance des autres cas possibles, avérés ou attendus. Selon Figari, même si les référents ne sont pas toujours explicites pour les évaluateurs eux-mêmes, ils sont toujours agissants. Dans un mouvement dialectique qui explique le fait que tout n'est pas joué d'avance, un cas évalué peut amener les évaluateurs à revoir leur représentation de tous les cas possibles comme de ceux qui sont attendus. Comme l'ont analysé Bonniol et Vial (1997), le modèle de référence est ajusté, dans une évaluation donnée, aux vues des « produits réels ». Les référents sont donc instables et évoluent en cours d'évaluation. L'évaluation peut donc produire non seulement de la connaissance sur un phénomène donné à évaluer, mais elle peut aussi potentiellement enrichir ou modifier la connaissance de tous les phénomènes de la même classe de situations.

On pourrait penser que le recours à une diversité de référents issus de l'expérience biaise l'évaluation, mais on observe plutôt que leur mobilisation contribue à l'activité et nourrit la démarche d'enquête collective qui se déploie. On observe ainsi une écoute effective entre membres du jury et un effet de l'évocation de ces références que l'on peut schématiser ainsi : lorsqu'un jury justifie un de ses jugements en évoquant et développant une référence d'expérience, cela suscite une réaction des autres membres du jury dans le sens d'un réexamen des jugements ou d'une relance des procédures d'examen du cas.

D'ailleurs, plus le jugement est difficile, plus les jurys recourent aux référents d'expérience. Les séquences dans lesquelles le recours à ces ressources est le plus intense lors d'un jury de VAE sont celles où l'interprétation du référentiel, la construction du jugement et du compromis collectif présentent des difficultés. On observe un phénomène plus spécifique aux situations d'interactions verbales à dimension argumentative : les participants puisent dans les ressources de leurs univers de référence pour justifier leurs jugements, appuyer des hypothèses ou les discuter. Cela consiste à introduire et relier des critères entre eux, identifier des indicateurs, présenter et discuter leur valeur. On peut penser que le recours à ces référents d'expérience est d'autant plus nécessaire que l'évaluation est plus indéterminée. Nous parlons d'indétermination pour désigner le fait que les situations d'évaluation ne peuvent être standardisées : pour un dossier de VAE, la nature de l'expérience du candidat et les choix de présentation et d'organisation de son dossier

en font un objet d'évaluation toujours particulier qui relève pourtant d'un même cadre référentiel. De même, pour les épreuves pratiques d'évaluation, il est impossible de proposer exactement la même situation à chacun des candidats : les conditions de réalisation des tâches varient, et ceci d'autant plus que les situations sont plus proches de situations de travail effectives. Ce qui rend le recours à l'expérience *produit* des membres du jury encore plus crucial.

L'EXPÉRIENCE DE L'EXPÉRIENCE DES PHÉNOMÈNES, OU DES ÉPISTÉMOLOGIES EN ACTES

Si les évaluateurs professionnels que nous avons rencontrés puisent largement dans leur expérience immédiate et passée des phénomènes pour évaluer (ainsi que dans celle des autres), ils s'en méfient aussi. Le recours à l'expérience s'accompagne, dans le même mouvement, d'une mise à distance de celle-ci. Les évaluateurs mettent en place les conditions pour assurer leurs jugements et, parfois, refréner leurs propres tendances à juger trop vite. En effet, les référents expérientiels mobilisés par les évaluateurs ne concernent pas uniquement les phénomènes évalués, mais également les manières d'évaluer, de porter des jugements, et donc de connaître et d'apprécier la valeur des phénomènes en jeu. Les évaluateurs mobilisent ainsi leur expérience d'autres situations d'évaluation, c'est-à-dire de situations où ils ont eu à faire l'expérience des phénomènes. Autrement dit, ils puisent dans leur expérience de l'expérience des phénomènes.

Une prévention des biais pour « bien évaluer »

Cela se manifeste, chez les évaluateurs que nous avons rencontrés, par des tentatives pour se prémunir contre des biais susceptibles de perturber la bonne marche de leur activité. Tous ont exprimé un souci de « bien évaluer » en étant attentifs à certaines propriétés des jugements produits (un souci d'équité, de justice, de crédibilité, de non remise en cause, que « ça tienne la route »...). Les préoccupations liées au fait d'assurer la légitimité des jugements produits (à travers la mobilisation de concepts pragmatiques comme l'équité, la validité, la crédibilité...) sont si centrales qu'elles constituent un des champs conceptuels qui orientent les activités de conception et de passation de l'évaluation, comme cela a été montré à propos des évaluateurs de programmes (Tourmen, 2007).

Ces préoccupations se traduisent par certains actes : les évaluateurs expriment et mettent en doute leurs « impressions » et premiers jugements, cherchent à poursuivre l'enquête quand elle ne les satisfait pas, recherchent l'accord des autres et suscitent une discussion en cas de désaccord, etc. Ils

utilisent aussi, pour ce faire, des instruments qui sont autant d'instruments psychologiques au sens où l'entend Vygotsky (1934/1997), médiatisant le rapport à sa propre activité de pensée : des grilles de critères qu'ils s'obligent à passer systématiquement en revue pour tous les candidats – comme dans le cas de l'évaluation en lycée agricole, où l'enseignante, en incitant le professionnel à reconsidérer l'ensemble des critères lors de chaque délibération, se pose en garante de l'équité de traitement des candidats –, des procédures de débat et de prise de parole dans le collectif qu'ils s'obligent à respecter, etc. Nous avons ainsi observé une gestion collective du degré de sévérité chez les jurys de VAE, le recours aux référentiels (métiers, formation...) donnant lieu à des échanges comparant candidats issus d'un parcours de formation et candidats à un diplôme par la voie de la VAE. On trouve ainsi explicitement mobilisés et utilisés pour l'action les concepts d'équité, de moyenne (« pour nos élèves on demande la moyenne, pas la perfection »), de compensation (« on ne peut pas demander à un candidat d'avoir tout appris, de tout connaître »), etc. Les évaluateurs comparent le mode d'évaluation lié à la VAE aux autres modes d'évaluation en cours dans les diplômes concernés et le degré de fiabilité des uns et des autres. Les jurys sont en effet en situation atypique quand ils agissent en situation de VAE et ils ne peuvent pas facilement mobiliser des formes d'action routinières. Ils sont davantage en proie aux doutes concernant la possibilité de bien évaluer via la VAE, d'être à la fois équitables et rigoureux. Ceci constitue un moteur pour le déploiement d'activités visant à prendre des précautions, à garantir le jugement et à se tirer honorablement de la délicate affaire que reste l'évaluation en VAE. On voit ainsi les jurys faire référence à toutes les autres situations d'évaluation comme à des cas d'école pour comparer, mais aussi, plus subtilement, comme à des précédents pouvant faire jurisprudence.

Ces préoccupations se retrouvent aussi dans le choix des méthodes de recueil de données, en évitant par exemple de chercher à observer des effets qui n'auraient pas eu le temps de se produire (Tourmen, 2007). Les évaluateurs réinvestissent ainsi dans leur rôle social une des lois de la conversation, mise à jour par les études en pragmatique linguistique : quand on prononce un énoncé assertif sur l'état du monde, on peut être tenu de le justifier. Ils déploient alors un certain nombre d'actes – allant de la constitution d'un référentiel au choix de méthodes de recueil de données en passant par des méthodes pour les interpréter – pour garantir, assurer leur jugement et en limiter les biais.

Le cadre du travail est, lui aussi, porteur d'une exigence de rigueur. Ainsi, les textes législatifs qui encadrent l'évaluation dans le système scolaire posent comme exigence « l'équité » des épreuves ; la charte de la Société française de l'évaluation³ pose comme principes « la prise en compte de la

3. Voir http://www.sfe-asso.fr/sfe-evaluation.php?menu_id=191

pluralité des points de vue », « l'impartialité » des évaluateurs, la « transparence » du processus, etc. Un certain nombre de procédures existent pour décourager une trop grande subjectivité des évaluateurs : le recours à des jurys mixtes (enseignants/professionnels), la présence de référentiels (certification, métiers, compétences...) devant servir de base commune à tous les évaluateurs, l'exigence de justifier les jugements au regard de critères préalablement définis, notamment.

Le cadre du travail des évaluateurs reste toutefois contradictoire dans le sens où, si plusieurs éléments concourent à encourager la rigueur, d'autres tendraient à la décourager. Le manque de temps et de moyens, voire le manque d'intérêt pour ce type d'épreuve chez certains enseignants, peuvent entraîner des stratégies d'évaluation *a minima* où les évaluateurs visent au plus rapide, au détriment d'une réflexion plus poussée sur les conditions de validité de leur évaluation. Comme cela a été montré chez des évaluateurs de politiques publiques, des préoccupations de « faisabilité » et d'« utilité » du travail rentrent souvent en conflit avec des préoccupations de validité des conclusions de l'évaluation, ceci donnant lieu à d'incessants compromis en situation (Tourmen, 2007).

Des épistémologies en actes ?

Ces actes – qui témoignent d'une prise en compte effective du rapport aux phénomènes et des risques de distorsion liés à leur étude et évaluation – ne sont pas tous impensés ou non conscients. Les évaluateurs que nous avons rencontrés ont ainsi développé, au fil du temps, des connaissances sur leur propre manière de connaître et de juger des phénomènes qu'il leur est demandé d'évaluer ainsi que sur les risques, les biais de l'évaluation et les moyens de les neutraliser, de les éviter. Ils expriment une confiance plus ou moins grande dans la justesse de leurs jugements ou de ceux de leurs collègues évaluateurs, dans leur capacité à discerner le réel sans erreur, dans leur capacité à ne pas être influencés par les évalués ni même par leurs propres préjugés (au sens littéral de croyances qui servent à juger avant). Les évaluateurs en lycée agricole ont ainsi tenté de limiter les biais de représentativité de leur épreuve en ayant conscience que cela reste « une situation simulée » (selon le professionnel, « c'est un peu dans le faux parce qu'on a des vaches qui sont faciles »). Ils se défendent contre certaines stratégies possibles des élèves pour tenter d'influencer leurs jugements (« c'est typiquement lui ça, est-ce que je peux pas gratter un point en disant ça ? [...] Il a réussi à me placer une réponse sur un tout autre sujet pour essayer de gagner des points »). Ils ont aussi témoigné d'une confiance dans la parole de l'autre évaluateur qui permet de limiter l'examen des jugements prononcés (selon l'enseignante, « on a la même vision du métier »). À un moment donné, l'enseignante demande même explicitement au professionnel : « t'as envie d'être gentil ? »

Nous trouvons trace ici de ce que nous appelons une « épistémologie en actes » des évaluateurs, que l'on pourrait définir comme un ensemble de représentations opératoires et non nécessairement conscientes ou verbalisées qui guident leur rapport à la certitude, aux objets évalués, à leurs propres processus psychiques de jugement, dans la lignée de ce qu'avaient supposé Shadish, Cook et Leviton (1991). Les évaluateurs se construisent une expérience sur l'expérience des phénomènes, qui influence leurs évaluations futures et qu'ils peuvent transmettre aux nouveaux venus.

Cet ensemble de représentations pour l'action – ou « d'invariants opératoires » – peut devenir l'objet d'une réflexivité. Allal et Mottier Lopez (2009) ont par exemple observé que certains enseignants pouvaient prendre conscience de leurs manières de noter et de juger. Merle affirme toutefois que tous les enseignants ne se posent pas cette question de façon explicite. « Faire part de sa "cuisine" évaluative et partager ses doutes nécessitent de dévoiler les limites de son propre jugement. Cette mise en cause professionnelle ne constitue pas une démarche ordinaire » (2007, p. 61), sauf lorsque le travail évaluatif est collectif comme dans les cas que nous avons observés. Cela peut ainsi devenir conscient. Certains évaluateurs vont-ils jusqu'à adopter, comme le leur suggère Patton (2008), des stratégies de doute conscientes pour lutter contre leurs tendances spontanées comme le « biais de confirmation » ? Différents évaluateurs mobilisent-ils différentes épistémologies en actes, selon leurs profils et leur formation, leurs statuts, leurs univers de travail et leurs préférences, comme le laisse supposer une étude comparative sur des évaluateurs de programmes (Fitzpatrick, 2004) ? S'inscrivent-ils dans différentes « épistémologies » selon les « paradigmes » d'évaluation auxquels ils se réfèrent, comme l'affirme Rodrigues (2001) ? Jusqu'où vont-ils dans « l'organisation de leur scepticisme », pour reprendre l'expression que Merton avait appliquée aux scientifiques (1973) ?

Des recherches resteraient à poursuivre sur ce point, même si l'étude des « épistémologies en actes » des évaluateurs demeure difficile et coûteuse en temps car elle nécessite d'en passer par l'analyse clinique de leur activité, toujours située, dans laquelle il s'agit de débusquer des traces de ce qu'ils tiennent pour vrai sur ce qu'il est convenu de tenir pour vrai, et de décrire des actes qu'ils réalisent sans toujours avoir conscience qu'ils révèlent leurs conceptions de leur rapport aux phénomènes.

CONCLUSION

Si la littérature en évaluation s'est construite sur une volonté de rupture avec les évaluations spontanées de la vie de tous les jours, l'activité d'évaluateurs au travail, telle que nous avons pu l'observer, n'en reste pas moins structurée par l'expérience des phénomènes évalués, expérience *processus* et

expérience *produit* qui nourrissent le processus d'évaluation et permettent de s'en distancier dans un même mouvement. L'enquête évaluative n'a pas lieu *ex nihilo* et dans l'absolu : elle est au contraire orientée et influencée par l'expérience que les uns et les autres ont des phénomènes, qu'elle influence en retour. L'expérience *produit* joue ainsi un rôle fondamental d'organisation de l'activité évaluative comme répertoire de cas et de situations rencontrées, ou, à un niveau plus conceptuel, comme système de représentations (plus ou moins conscientes) des caractéristiques estimées pertinentes pour l'action et de connaissances ou croyances sur les buts à atteindre, les résultats à obtenir, les objets et les phénomènes en jeu.

Les évaluateurs que nous avons rencontrés mettent toutefois en place différents actes destinés à favoriser une prise de distance avec les jugements issus de leur expérience des phénomènes, à les médiatiser par le doute, l'enquête rigoureuse, la discussion, divers instruments (référentiels, grilles de critères...) qui fonctionnent comme autant d'instruments psychologiques (Vygotsky, 1934/1997) pour se prémunir contre les tendances de l'esprit à juger vite et sur la base d'informations partielles. Il semble toutefois, selon d'autres études, que tous les évaluateurs n'aient pas le même degré de conscience ni de prise en compte de ces phénomènes et que certains restent davantage sous l'influence de leur expérience immédiate, même si nous n'avons pu l'observer sur nos terrains. Certes, l'exercice demeure périlleux. Mais que serait une connaissance des phénomènes humains qui ne s'appuierait sur aucune intuition, sur aucun savoir antérieur ? Une connaissance coupée de l'expérience est-elle possible ?

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abernot, Y. (1988). *Les méthodes d'évaluation scolaire*. Paris : Bordas.
- Allal, L. & Mottier Lopez, L. (2009). Au cœur du jugement professionnel en évaluation : des démarches de triangulation. *Les dossiers des sciences de l'éducation*, 22, 25-40.
- Bachelard, G. (1971). *La formation de l'esprit scientifique*. Paris : Vrin.
- Barbier, J.-M. (1985). *L'évaluation en formation*. Paris : Presses Universitaires de France.
- Barreau, H. (1992). *L'épistémologie* (Que sais-je ? N° 1475). Paris : Presses Universitaires de France.
- Bonniol, J.-J. & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles : De Boeck.
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris : Presses Universitaires de France.
- Cardinet, J. (2003). Cinq dispositifs pour vérifier le progrès. *Mesure et évaluation en éducation*, 26(1-2), 51-59.

- Dewey, J. (1993). *Logique, théorie de l'enquête*. Paris : Presses Universitaires de France.
- Figari, G. (1994). *Évaluer : quel référentiel ?* Bruxelles : De Boeck.
- Figari, G. (2006). Le référentiel entre théorie et méthodologie. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 101-108). Paris : L'Harmattan.
- Fitzpatrick, J. L. (2004). Exemplars as case studies : Reflections on the link between theory, practice and context. *American Journal of Evaluation*, 25, 541-559.
- Lemieux, V. (2002). *L'étude des politiques publiques*. Québec : Les Presses de l'Université Laval.
- Leplat, J. (1997). *Regards sur l'activité en situation de travail. Contribution à la psychologie ergonomique*. Paris : Presses Universitaires de France.
- Martin, M. (2011). *Étudier les normes de recrutement liées à l'activité du BTP dans une démarche de préconisation de formation aux outils de Techniques de Recherche d'Emploi*. Travail d'étude de licence professionnelle, Université de Bourgogne, Dijon.
- Mayen, P. (2009). Expérience et formation des adultes. In J.-M. Barbier, E. Bourgeois, J. C. Ruano-Borbalan & G. Chapelle (Éd.), *Encyclopédie de la formation* (pp. 763-780). Paris : Presses Universitaires de France.
- Mayen, P., Mayeux, C. & Savoyant, A. (2006). Construire la référence : une activité permanente des jurys de VAE. In G. Figari, P. Rodrigues, M. P. Alves & P. Valois (Éd.), *Évaluation des compétences et apprentissages expérientiels* (pp. 126-143). Lisbonne : ADMEE-Europe et Educa.
- Mayen, P. & Tourmen, C. (2009). Les jurys de VAE ont-ils quelque chose à apprendre ? In A. Savoyant & P. Mayen (Éd.), *Élaboration et réduction de l'expérience dans la validation des acquis de l'expérience* (pp. 59-69). Marseille : Relief28, Cereq Relief.
- Merle, P. (2007). *Les notes. Secrets de fabrication*. Paris : Presses Universitaires de France.
- Merton, R. K. (1973). *The Sociology of Science : Theoretical and Empirical Investigations*. Chicago : University of Chicago Press.
- Nadeau, M. A. (1988). *L'évaluation de programme, théorie et pratique*. Québec : Les Presses de l'Université Laval.
- Pastré, P., Mayen, P. & Vergnaud, G. (2006). Note de synthèse sur la didactique professionnelle. *Revue française de pédagogie*, 154, 145-198.
- Patton, M. Q. (2008). *Utilization-focused Evaluation* (4^e éd.). Thousand Oaks : Sage publications.
- Rodrigues, P. (2001). L'entretien de recherche comme pratique d'évaluation. In G. Figari & M. Achouche (Éd.), *L'activité évaluative réinterrogée : regards scolaires et socioprofessionnels* (pp. 219-225). Bruxelles : De Boeck.
- Shadish, W. R., Cook, T. D. & Leviton, L. C. (1991). *Foundations of Program Evaluation Theories of Practice*. Newbury Park : Sage publications.

- Tourmen, C. (2007). *Les compétences des évaluateurs, le cas des évaluateurs de politiques publiques*. Thèse de doctorat en Sciences de l'éducation, Université de Grenoble II.
- Tourmen, C. (2009). L'activité évaluative et la construction progressive du jugement. *Les dossiers des sciences de l'éducation*, 22, 101-119.
- Tourmen, C. & Deblay, S. (2011). Évaluer des compétences : une situation à enjeux. In M. Vantourout, R. Goasdoué & L. Mottier Lopez (Éd.), *L'évaluation des compétences à l'épreuve du réel*. Manuscrit soumis pour publication.
- Vergnaud, G. (1992). Qu'est-ce que la didactique ? En quoi peut-elle intéresser la formation des adultes peu qualifiés ? *Éducation permanente*, 111, 19-32.
- Vygotsky, L. S. (1934/1997). *Pensée et langage*. Paris : Éditions Sociales.

**LES ACTIVITÉS D'ÉVALUATION :
QUELLES MODÉLISATIONS
MISES EN ŒUVRE ?**

L'évaluation des élèves. Une modélisation interactionniste des pratiques professorales

**Pierre Merle
IUFM de Bretagne
et Université européenne de Bretagne**

L'analyse sociologique est divisée originellement en deux courants fondateurs. L'école française est centrée sur l'analyse des faits sociaux, « manières d'agir, de penser et de sentir, extérieurs à l'individu et qui [ces faits sociaux] sont doués d'un pouvoir de coercition en vertu duquel ils s'imposent à lui » (Durkheim, 1895/1986, p. 5). L'école allemande se donne pour objet de comprendre par interprétation l'activité sociale (Weber, 1921/1995). Elle est à l'origine du courant interactionniste. Cette division originelle de la pensée sociologique ne rend compte que de façon simplifiée de ces deux paradigmes : la question de l'individu, notamment celle du lien entre « conscience collective » et « conscience individuelle », est paradoxalement présente chez Durkheim (1893/1994), et la sociologie wébérienne a su intégrer dans ses analyses de l'activité individuelle l'effet du contexte historique, en l'occurrence l'éthique protestante (Weber, 1964). Cette division originelle de la sociologie, même si elle reste structurante des recherches actuelles, simplifie aussi la diversité contemporaine de la sociologie. Pour la France, Ansart (1990) distingue plusieurs orientations spécifiques. L'analyse bourdieusienne se situe notamment dans la postérité de Durkheim et du structuralisme et adhère au principe d'objectivation statistique du social. La sociologie de Boudon (1984) est centrée sur l'analyse des comportements des acteurs individuels et s'oppose au « préjugé structuraliste ».

Toutefois, dans le champ éducatif, les deux conceptions sociologiques originelles – le déterminisme *versus* l'interactionnisme – demeurent une

clé d'interprétation pertinente tant ces orientations amènent à construire de façon différente, voire opposée, un certain nombre d'objets et tout particulièrement les pratiques d'évaluation et de notation des compétences des élèves. Ces pratiques constituent un fait social, au sens durkheimien du terme, en raison de leur forte institutionnalisation, leur pérennité historique, leur rôle central dans l'orientation des populations scolaires dans les différentes filières, leurs implications pour l'avenir des élèves et leur diffusion spatiale. Ces pratiques sont en effet présentes, avec des ajustements nationaux, dans la quasi-totalité des systèmes éducatifs.

La centralité des pratiques de notation explique l'abondance considérable des recherches qui y sont consacrées. Historiquement, ces pratiques de notation ont fait l'objet de trois approches qui sont conceptuellement distinctes. La première – les études docimologiques – a poursuivi une orientation descriptive, de type statistique, non des pratiques d'évaluation sommative, mais de leurs conséquences : les notes. Le contexte de la classe est absent. L'approche des docimologues est équivalente à la mesure de la conformité d'un produit de consommation : quelles sont ses spécificités techniques ? Quelles sont les marges d'erreur standard ? L'approche docimologique est a-historique et a-sociale. Les deux autres analyses des pratiques de notation des professeurs – le modèle déterministe et le modèle interactionniste – ont poursuivi une tout autre ambition, celle de proposer un cadre d'interprétation aux variations des notes en rattachant celles-ci à des contextes sociaux : les rapports sociaux de sexe, la structure de classe, les contraintes pédagogiques du professeur, la complexité des relations maître-élèves... En ce sens, contrairement à l'approche docimologique enfermée dans le paradigme statistique de la mesure, les analyses déterministe et interactionniste sont des modélisations : elles construisent des schémas d'intelligibilité.

Pour l'approche docimologique et les modélisations déterministe et interactionniste, sont présentés ci-après quelques raisons de leur émergence, les principes qui les fondent et les principaux résultats de la recherche. L'accent sera mis sur le modèle interactionniste pour trois raisons. Il s'agit de la modélisation la moins connue ; elle permet en partie de rendre compte des connaissances cumulées par les modélisations antérieures ; elle a l'avantage d'intégrer l'expérience subjective des professeurs confrontés à la double contrainte institutionnelle de faire progresser leurs élèves et de les noter.

L'APPROCHE DOCIMOLOGIQUE

La docimologie – du grec *dokimé* qui signifie « épreuve » et de *logos*, « science » – constitue historiquement la première recherche de type scientifique sur la notation. Elle consiste en l'étude statistique des variations des notes attribuées à une même copie. La justification de la première grande recherche docimologique, réalisée sur les notes au baccalauréat, est apportée

par Laugier et Weinberg (1936) : « le baccalauréat occupe en France une place prépondérante comme régulateur de l'accès aux professions libérales » (p. 71). À l'époque, moins de 3 % d'une génération est titulaire du baccalauréat et Laugier et Weinberg s'interrogent sur la fiabilité et l'équité de la procédure de sélection des élites scolaires et sociales. Cette préoccupation, centrale dans l'institution scolaire, demeure toujours présente en ce début du XXI^e siècle.

Un des résultats majeurs des recherches de Laugier et Weinberg a été de montrer l'incertitude de la notation au baccalauréat. À partir d'échantillons représentatifs de lots de 100 copies rédigées dans six disciplines ayant fait l'objet d'un écrit au baccalauréat, l'enquête a consisté à soumettre celles-ci à cinq correcteurs. L'étude a eu pour objet de calculer les écarts de notes entre les six correcteurs (cinq corrections par copie auxquelles il faut ajouter la correction réalisée par le correcteur du baccalauréat) et a notamment donné lieu, par discipline, au calcul de l'écart maximum entre les six notes par copie et des écarts revenant le plus fréquemment. L'écart maximum de notation, pour une même copie, est considérable pour les disciplines de type littéraire telles que la philosophie (12 points d'écart) ou la composition française (13 points). Plus surprenant encore est l'écart maximum, pour la même copie, pour les disciplines scientifiques. Il est de 9 points pour les mathématiques, 8 points pour la physique. L'écart le plus fréquent, indicateur plus pertinent, est également important. Il est de 7 points en philosophie et composition française ; de 4 points en physique et en mathématiques.

Ces données montrent l'incertitude considérable attachée à la correction des copies quelle que soit la discipline considérée. Laugier et Weinberg (1936) en avaient conclu que pour obtenir la « note vraie », il fallait recourir à la moyenne de 13 correcteurs en mathématiques, 78 en composition française, 127 en philosophie... La fiabilité incertaine de la correction des copies dans les disciplines scientifiques et littéraires pose sans aucun doute la question de la pertinence du jugement professoral. Dans la recherche menée par Laugier et Weinberg, l'évaluation professorale relative à l'admission ou à l'ajournement des candidats est, en effet, loin d'être consensuelle : une grande part des candidats est soit refusée, soit admise, selon l'examineur qui assure la correction de leurs copies. Ainsi en français, seulement 21 % des candidats sont refusés unanimement (note inférieure à 10 pour les six correcteurs), 9 % seulement des candidats sont reçus par l'ensemble des correcteurs, et 70 % sont refusés ou admis selon le correcteur. En mathématiques, discipline où l'accord des correcteurs est le plus fréquent, 36 % des candidats sont pourtant refusés ou admissibles selon l'examineur.

La recherche de Laugier et Weinberg (1936) est à l'origine d'un grand nombre d'études. Celles de Noizet et Caverni (1978), relatives aux épreuves du baccalauréat, ont confirmé les résultats de Laugier et Weinberg. Des études plus ponctuelles sont également riches d'enseignement. Aymes (1979) a soumis huit copies de baccalauréat à la notation de 10 correcteurs des épreuves

du bac. Chaque correcteur disposait du barème détaillé utilisé par tous les correcteurs de l'épreuve. Première constatation, l'écart entre les notes attribuées sur la même copie par les 10 correcteurs de mathématiques est loin d'être négligeable alors même que le nombre de correcteurs est relativement limité. Pour une même copie, la note la plus élevée est de 16.5/20, la plus faible est de 12/20. Deuxième constatation, l'utilisation d'un barème de notation au point près, voire au demi-point près, ne constitue pas une garantie de précision de la correction. Les écarts de notation peuvent être nuls ou très limités entre les correcteurs sur des questions notées sur deux ou trois points. Inversement, une grande incertitude peut caractériser la notation d'une question notée sur un point. La finesse d'un barème ne permet des notations plus concordantes que s'il existe une définition précise, entre les différents correcteurs, sur ce qui est attendu exactement pour telle ou telle question, que celle-ci soit notée sur trois points ou un seul. Un tel constat indique qu'il ne faut pas attendre une grande efficacité de l'introduction d'un barème dans la correction d'une dissertation en dehors d'une socialisation longue des correcteurs à des exigences effectivement communes. La discussion sur un barème constitue évidemment une des modalités de cette socialisation évaluative. La recherche la plus récente (Suchaut, 2008), portant sur les copies de baccalauréat en sciences économiques et sociales, débouche sur le même résultat que ceux trouvés par Laugier et Weinberg 70 ans plus tôt. Les écarts maximums et moyens de notation pour chaque copie sont considérables.

Le modèle docimologique prémunit contre l'idée d'une note vraie, utopie scolaire et, souvent, croyance professorale. L'approche docimologique a été, et demeure, d'un grand intérêt pour trois raisons essentielles. D'abord, elle montre que la notation d'une copie est un travail d'expert et, à ce titre, elle impose toujours un travail d'interprétation qui ne se répète jamais à l'identique. Pour cette raison, la notation d'une copie doit toujours être comprise comme une mesure approximative de la compétence de l'élève. Ensuite, l'approche docimologique montre que l'évaluation des compétences des élèves est assurée de façon plus équitable par la multiplication des mesures, ce qui est le cas lors des épreuves du baccalauréat. Enfin, elle montre, indirectement, l'intérêt des mesures standardisées des connaissances afin de pouvoir établir des comparaisons des compétences des élèves entre classes, entre établissements et dans le temps. Une limite essentielle de cette approche docimologique tient toutefois à la non prise en compte du contexte social des pratiques professorales.

LE MODÈLE DÉTERMINISTE

Le modèle déterministe est une façon de désigner un ensemble de recherches qui se sont focalisées, non sur l'aléa intrinsèque de la notation d'une copie, mais sur les *déterminants contextuels* liés aux pratiques professorales

d'évaluation. L'hypothèse au fondement de ces analyses est de considérer que l'évaluateur ne note jamais une copie en soi. La notation est un travail de comparaison des travaux d'élèves qui ne peut pas s'exercer indépendamment d'informations scolaires et extrascolaires spécifiques. Ces recherches ont pour objet d'isoler l'effet de variables contextuelles susceptibles de déterminer ou de biaiser les pratiques de notation des correcteurs.

L'étude de l'effet de l'ordre de correction des copies, spécifiquement étudié par Bonniol (cité par Noizet & Caverni, 1978), constitue une première analyse des déterminants de la notation. D'une part, les copies placées dans le premier tiers du paquet sont en moyenne notées de façon plus indulgente (corrigées dans l'ordre inverse, ces copies sont notées plus sévèrement). D'autre part, chaque copie est notée différemment selon sa position par rapport aux autres copies : la même copie est surévaluée lorsqu'elle vient après une copie faible et sous-évaluée lorsqu'elle vient après une copie forte. Enfin, si trois copies médiocres sont placées en début du paquet, la moyenne générale s'élève. Le phénomène inverse est observé quand trois bonnes copies remplacent les trois mauvaises. Le barème n'est pas un décompte mécanique. Il résulte d'une interprétation des réponses des élèves et ses modalités concrètes d'application sont guidées par une norme scolaire implicite : parvenir à une moyenne ni trop élevée, ni trop basse.

Cette recherche apporte des connaissances sur l'incertitude de l'évaluation des copies dans des situations anonymes de correction. Or, dans le cadre de la classe, la notation des élèves a la particularité de se réaliser dans une situation où l'anonymat des élèves n'est ordinairement pas la règle : les professeurs demandent en effet généralement à leurs élèves de remplir des fiches de renseignements sur lesquelles figurent des informations telles que la profession des parents, le cursus scolaire, les notes obtenues l'année précédente, l'âge... (Merle, 2007). En dehors d'une telle fiche, les professeurs peuvent de toute façon être amenés à connaître une partie de ces informations. Quelles incidences ces informations scolaires et extrascolaires peuvent-elles avoir sur l'évaluation des compétences scolaires des élèves ? Les chercheurs ont répondu à cette question en créant des situations fictives de correction qui permettent d'isoler l'effet de la possession de telles ou telles informations sur la notation des devoirs des élèves.

Caverni, Fabre et Noizet (cités par Noizet & Caverni, 1978) ont ainsi isolé l'effet sur la notation d'un devoir de la connaissance par le correcteur du niveau scolaire de l'élève. Dans une première étape, l'expérience a consisté à faire corriger par plusieurs correcteurs 12 copies d'anglais de baccalauréat de niveau très différent (ces copies sont dactylographiées). Dans une seconde phase (ignorée initialement par les correcteurs), six copies, également dactylographiées, mais de niveau cette fois à peu près équivalent, ont été soumises à la correction des mêmes professeurs. Sur chaque copie figurait une note attribuée antérieurement par le correcteur

et censée avoir été obtenue par l'élève (ces notes apposées sur les copies étaient soit fortes soit faibles). L'association entre la copie à corriger et la note initialement mise était fictive, mais une telle association a orienté significativement l'évaluation des enseignants : les mêmes copies sont notées de façon sensiblement différente selon que la note antérieure, attribuée fictivement, est faible ou forte. Dans cette seconde phase de l'expérience, le correcteur est influencé, lors de l'évaluation de chaque copie, par l'évaluation qu'il a réalisée initialement. Les auteurs de cette recherche concluent sur ces données en indiquant que le modèle de référence de l'évaluateur est constitué pour une part déterminante, par l'évaluation précédemment produite par l'évaluateur lui-même.

Outre le niveau de l'élève, son statut scolaire influence également l'évaluation des performances des écrits. Il y a déjà plus d'un quart de siècle, Bonniol, Caverni et Noizet (cités par Noizet & Caverni, 1978) ont fait procéder à la correction de huit copies provenant pour moitié de 6^e de type 1 (réputées) et de 6^e de type 3 (peu réputées). L'étude aboutit à un résultat qui concorde avec les données précédentes : lorsque les copies sont attribuées à des élèves de 6^e de type 1, elles sont sensiblement mieux notées que lorsqu'elles sont attribuées à des élèves de 6^e de type 3. Dans le premier cas, les copies obtiennent plus de 10/20 alors que dans le second cas elles n'obtiennent pas la moyenne. Ces deux types d'études portant sur l'effet de la connaissance du niveau et du statut scolaires sur la notation montrent, de façon manifeste, un phénomène de dépendance entre évaluations : toutes informations qui d'une manière ou d'une autre sont dissonantes, entraînent de la part d'un correcteur un effort pour mieux les faire s'accorder.

En début d'année, les professeurs font souvent remplir à leurs élèves une fiche sur laquelle ceux-ci indiquent généralement la profession de leurs parents. C'est pour cette raison qu'il est utile de connaître les effets de cette information sur l'élaboration du jugement du professeur. Dans cette perspective, Weiss (cité par de Landsheere, 1976) a sélectionné deux compositions françaises rédigées par des élèves de 4^e. Celles-ci ont été dactylographiées et soumises à la correction de deux groupes de 46 enseignants. Les deux copies ont été transmises au premier groupe de correcteurs en mentionnant que la première copie était rédigée par un enfant doué, fils d'un rédacteur d'un quotidien connu alors que la seconde copie était rédigée par un élève moyen qui aime lire des bandes dessinées et dont le père et la mère sont employés. Au second groupe de correcteurs, les informations extrascolaires associées à chaque copie étaient inversées. Les enseignants avaient, pour chaque copie, à noter séparément l'orthographe, le style, le fond et l'ensemble de la copie. Pour les quatre aspects considérés, la copie qui bénéficie d'un préjugé favorable est significativement mieux notée.

La recherche de Weiss présente toutefois l'inconvénient de ne pas distinguer les effets des deux variables introduites dans l'expérience (statut

scolaire de l'élève et origine sociale). Une étude de Pourtois (1978) a permis de dépasser les limites méthodologiques du travail de Weiss. L'expérience a consisté à faire corriger par plusieurs correcteurs quatre compositions de français (classe de 6^e) associées fictivement à deux types d'informations : « copies des enfants issus de familles socialement défavorisées » ou « copies des enfants issus de familles socialement favorisées », informations inversées pour la moitié des correcteurs. Les copies bénéficiant d'un préjugé laudatif font l'objet d'une notation plus clémente. L'écart le plus sensible de notation ne porte pas sur le fond mais sur l'orthographe pour laquelle la définition de la faute n'est a priori pas subjective...

Ces recherches montrent que les procédures ordinaires de jugement des élèves dans le quotidien de la classe sont empreintes de « biais » ou d'erreurs systématiques d'évaluation produits par la détention d'informations scolaires et extrascolaires sur l'élève. En ce sens, les informations demandées par les professeurs dans les fiches de renseignements qu'ils font remplir par leurs élèves en début d'année favorisent ces biais d'évaluation. En termes d'action sociale, ces recherches montrent l'intérêt des échanges de copies entre professeurs au cours desquels les élèves bénéficient de corrections anonymes. Davantage pratiqués, ces échanges de copies favoriseraient l'harmonisation des critères d'évaluation des professeurs dont on sait qu'elle permet de diminuer les écarts d'appréciation selon le correcteur (Noizet & Caverni, 1978). Ces études éclairent aussi les débats relatifs à l'intérêt des examens ayant recours à l'anonymat des candidats. La thèse d'Oget (1999) a montré que si l'obtention du baccalauréat était réalisée uniquement par le contrôle continu, les redoublants, les garçons et les enfants d'origine populaire, moins bien notés à compétences égales pendant l'année, seraient moins souvent bacheliers.

L'intérêt de ces recherches est indéniable. Il faut regretter d'ailleurs qu'elles demeurent relativement peu connues alors même qu'elles permettraient aux professeurs de mieux comprendre les incertitudes attachées à leurs pratiques de notation. Cependant, ces résultats de recherche, établis à partir de situations fictives de correction, ne permettent pas de connaître les incertitudes effectives d'évaluation dans le quotidien de la classe. Sur ce point, les travaux de Duru-Bellat et Mingat (1993) ont apporté, en comparant moyennes annuelles en classe et résultats à des tests standardisés de compétence, plusieurs précisions. Les élèves « en retard », mais non redoublants, obtiennent un point de moins par année d'âge supplémentaire à niveau de connaissances identiques. Les élèves redoublants font également l'objet d'une notation plus sévère à compétence donnée. Les enfants de cadres supérieurs se distinguent des enfants des autres groupes sociaux : pour un niveau donné de résultats aux tests, ils obtiennent en moyenne un demi-point de plus que les enfants des autres milieux. Ce demi-point, biais social d'évaluation, représente presque le quart de la différence moyenne qui sépare les enfants de cadres des enfants d'ouvriers. Enfin, notées par leurs professeurs, les filles obtiennent en moyenne des notes supérieures de

0.78 point à celles des garçons à compétences identiques aux tests. Aucun biais social d'évaluation au détriment des élèves d'origine étrangère n'a été constaté, ce qui peut s'expliquer par la connaissance qu'ont les correcteurs de l'existence de préjugés raciaux et, en conséquence, par une vigilance accrue de leur part lors de la correction des copies d'élèves d'origine étrangère.

L'approche déterministe présente toutefois une limite. Elle postule l'existence des stéréotypes professoraux susceptibles d'expliquer la partialité de leurs notes. Au cours de la correction, le professeur devient particulièrement attentif à ce à quoi il s'attend si bien que ses attentes sont, pour une part, confirmées. Il s'agit d'une illustration des « effets d'attentes », de « prophéties autoréalisatrices », mis en évidence par Rosenthal et Jacobson (1971) à la fin des années soixante. Autrement dit, les professeurs, « victimes » ou « complices » de préjugés sociaux, favoriseraient, par exemple, les élèves issus de milieux aisés ou les bons élèves. Une telle hypothèse – celle de professeurs incapables de comprendre leurs propres pratiques – ne va pas de soi. Elle néglige aussi une donnée indissociable de l'activité ordinaire d'évaluation. Dans le quotidien, la notation de la classe n'assure pas seulement une fonction d'évaluation des élèves.

LE MODÈLE INTERACTIONNISTE

Les limites d'une interprétation des biais de l'évaluation scolaire en termes de préjugés sociaux – ceux-ci n'expliquent d'ailleurs qu'une partie limitée de l'incertitude de la notation – justifient le développement d'une théorie de l'évaluation fondée sur l'observation des interactions à l'intérieur de la classe (Merle, 2007). Cette orientation théorique émerge en raison des modalités d'imposition de l'autorité professorale, question centrale puisque l'autorité professorale est une condition indispensable à l'exercice des missions du maître : « faire cours » et permettre les apprentissages des élèves. Or, de nombreuses données indiquent que cette autorité connaît une perte d'influence certaine. Actuellement, l'autorité pédagogique ne peut, dans le cadre d'une relation personnalisée en partie routinière, s'appuyer durablement ni sur le charisme du professeur, ni sur l'autorité réglementaire réduite dont il dispose, ni enfin, sur un respect traditionnel de l'institution scolaire auprès d'un public scolaire pour lequel la scolarisation prolongée est souvent une nouveauté familiale.

Différents types d'arrangements évaluatifs

Dans cette perspective de recherche, la notation a évidemment toujours pour objet de classer les élèves selon leurs performances scolaires, mais les incertitudes associées à la note sont expliquées par l'existence d'*arrangements*

qui tiennent au contexte scolaire d'évaluation appréhendé à trois niveaux qui sont définis par l'établissement, les relations maître-élèves à l'intérieur de la classe et la personne du maître. Ces trois niveaux d'arrangements sont distingués pour des raisons heuristiques ; ils sont en fait en forte interdépendance. Ces arrangements contribuent à expliquer les biais sociaux de notation sans recourir à l'hypothèse de stéréotypes évaluatifs. Ils permettent aussi de concilier l'exigence institutionnelle de la notation avec la nécessité, parfois contraire, de permettre les apprentissages des élèves.

Le premier type d'arrangements concerne l'établissement et correspond à une contrainte externe pour le professeur. Dans une recherche précitée réalisée sur 17 établissements et 2500 élèves, Duru-Bellat et Mingat (1993) ont montré que, contrairement aux attentes, les notations les plus indulgentes sont attribuées plutôt dans les établissements où les élèves ont obtenu des résultats faibles aux tests de compétence, et inversement. Des approches ethnographiques (par exemple Barrère, 2002) aboutissent à des conclusions du même ordre. Ainsi dans les « collèges difficiles », une partie des enseignants sont amenés à des adaptations sensibles de leurs pratiques ordinaires d'évaluation, en évitant des contrôles qui aboutiraient à des notes jugées très basses, ou par exemple en adoptant une notation de la participation orale qui a pour seul objectif d'augmenter la moyenne des élèves.

Un second type d'arrangements correspond à une contrainte interne pour le professeur et concerne la classe. Des arrangements de ce type se réalisent lorsque le professeur, prenant en considération la bonne volonté de « ses » élèves, dans telle classe, décide de supprimer de la moyenne trimestrielle les notes d'un contrôle peu réussi, ou d'ajouter un devoir « facile » en fin de trimestre, etc. Dans la situation inverse – agitation, travail non fait, chahut – le professeur peut avoir recours à une « interrogation surprise » ou donner un devoir « difficile », sorte de sanction pour montrer aux élèves les conséquences de leur manque d'attention en cours. Ces arrangements concernent également les élèves considérés individuellement. Outre l'octroi d'une note de participation en cours, l'élève qui accepte de faire un travail supplémentaire, un exposé par exemple, ou de refaire un devoir raté, pourra bénéficier d'une note supplémentaire ou d'une note se substituant à ce devoir, ou bien encore d'une note qui ne sera prise en compte dans la moyenne que si elle dépasse 10/20, etc. (Merle, 2007). La dispersion des comportements des professeurs est grande dans ce domaine. Ceux-ci peuvent être, en effet, plus ou moins sensibles au sentiment d'iniquité que les élèves faibles ressentent à l'égard de leurs notes lorsque ces derniers ont le sentiment que la récompense que constitue la note obtenue n'est pas à la hauteur des efforts fournis. Dans ce type de situation particulière, le professeur peut être amené, au nom d'un équilibre nécessaire entre travail et gratification scolaires, à noter davantage les progrès réalisés par l'élève que le niveau atteint et normalement visé à tel ou tel niveau de scolarité.

Les arrangements individuels qui engagent de façon personnelle élève et maître sont indissociables, pour le maître, d'un troisième type d'arrangements. Il prend la forme d'une contrainte par rapport à soi : la notation du professeur est orientée par sa propre histoire scolaire et par les diverses significations que celui-ci associe à son activité de notation (évaluer de façon impartiale, aider, récompenser, sanctionner...). Une part des professeurs a une certaine conscience de ces arrangements et c'est notamment pour cette raison qu'ils sont favorables au maintien du bac qui constitue, dans leurs propos, une garantie d'équité scolaire.

UN EXEMPLE DE NÉGOCIATION INDIVIDUELLE DE LA NOTE

Question : Existe-t-il des différences entre la moyenne annuelle des élèves et les notes qu'ils obtiennent au baccalauréat ?

« Ça peut. Ah oui, dans les deux sens. Dans l'année par exemple, je peux avoir mis des notes convenables et pûuu (bruitage d'un dégonflement). Par exemple, je prends l'exemple de Bruno l'année dernière. Je l'avais en classe de première. Voilà le cas d'un élève qui ne maîtrise pas du tout l'écrit et n'arrive pas à structurer sa pensée [...]. Au début de l'année, première note, je lui mets 04/20. Il me fait un résumé 04. Une semaine après, il vient me voir comme ça, j'étais dans une classe, je me souviens très bien, et il passe dans le couloir, la porte était ouverte. Il en profite pour rentrer, pour m'interpeller : "comment ça se fait que vous m'avez mis 04 à mon devoir, et l'année dernière, j'avais un professeur de lettres qui me mettait douze, douze !" »

Je lui ai dit : "c'est fort possible. Qui c'est?... C'est elle ? (rires). Ah oui, ah bon, non, ça ne m'étonne pas du tout (rires). Bon, tu as eu douze l'année dernière, mais cette année, encore faudra-t-il le prouver. Tu as eu 04 au résumé, tu ne sais pas faire un résumé, si tu me montres que tu sais faire un résumé, je te mettrai douze et plus même. Je suis tout à fait disposée... mais c'est à toi de me fournir les preuves. Cette année, l'objectif c'est de te fournir une notation bac." N'empêche que Bruno s'est offert 04 à l'écrit du bac (rires). Bien. Et c'est tout à fait mérité.

Mais pendant l'année, euh... j'apprécie chez lui une culture que la majorité des élèves de la classe n'avait pas, des références à des lectures même si ce ne sont pas forcément les miennes, c'est quelqu'un qui lisait... Il y a eu donc des affinités qui se sont révélées, c'est quelqu'un qui intervenait facilement à l'oral. En plus, c'est l'exemple même de l'élève capable de faire de l'absentéisme plus ou moins chronique, donc pour lui donner envie de rester en cours, j'étais obligée d'envisager une notation thérapeutique. C'est le cas même de celui à qui j'ai offert une notation thérapeutique. Mais en étant clair avec lui et net, je veux dire : je te mets 09, je te mets 10, mais je te préviens que le jour du bac, c'est pas ça que tu auras » (professeur de français). (Merle, 2007, pp. 92-93)

Ces arrangements évaluatifs sont d'une grande variété et sont d'autant plus fréquents que l'enseignant est confronté non seulement à la question de l'ordre scolaire, à la gestion des relations maître-élèves dans la classe, mais simultanément à la question de la transmission du savoir et des rythmes d'apprentissage. Cette contrainte de l'action enseignante explique que les arrangements évaluatifs intègrent inévitablement une dimension didactique. La notation au demi-point près ne doit pas être comprise dans le cadre de la « note vraie » calculée par Laugier et Weinberg mais d'abord comme le résultat d'une « transaction » ou d'un « contrat » de type didactique (Chevallard, 1991). Le 9.5/20 n'exprime pas tant la précision de la mesure des performances qu'une sorte d'avertissement symbolique dont l'objet est de signaler à l'élève que celui-ci ne remplit pas totalement les exigences attendues spécifiques à sa classe et son établissement. Inversement le 10/20, attribué à un élève « faible » « qui s'accroche », a pour objet de lui montrer que le travail « peut payer » et qu'il lui faut « maintenir son effort ». Cette « notation thérapeutique » indique que, pour le professeur, les contraintes de l'apprentissage en termes d'estime de soi de l'élève et de sentiment d'efficacité l'emportent sur la contrainte institutionnelle du classement. Ce 10/20 est parfois le produit d'une sorte de négociation plus ou moins explicite de la note.

Arrangements et biais sociaux d'évaluation

Dans cette analyse interactionniste des pratiques d'évaluation, les notes scolaires deviennent un motif de la mobilisation des élèves, un moyen de paix scolaire dans et hors de la classe, et aussi, une dimension majeure de l'autorité du professeur qui dispose, plus ou moins consciemment, d'une certaine marge de manœuvre lorsqu'il accorde ou non une reconnaissance scolaire par la note. Cette analyse offre un cadre théorique pertinent à l'analyse de l'incertitude de la notation et des biais sociaux d'évaluation précédemment présentés. La mise en parallèle des comportements en classe des élèves et des évaluations de leurs écrits permet en effet de comprendre, au moins partiellement, les biais d'évaluation constatés. Ainsi les comportements scolaires des filles, davantage conformes aux attentes professorales (Duru-Bellat, 1995), expliqueraient leur surévaluation scolaire à compétences équivalentes aux garçons. Et le comportement plus agité, voire contestataire, des garçons d'origine populaire (Felouzis, 1994) expliquerait leur notation en moyenne un peu plus sévère. Inversement, les élèves d'origine aisée, dont on sait qu'ils ont une meilleure maîtrise du métier d'élève et qu'ils prennent plus facilement la parole en classe (Felouzis, 1994 ; Sirota, 1988) bénéficient d'arrangements individuels en leur faveur. Par leur rôle de « locomotive », ils aident le professeur à faire cours. L'analyse des entretiens menés auprès des enseignants montre d'ailleurs que dans le quotidien de la classe se construisent des images scolaires d'élèves plus ou moins favorables à cette

reconnaissance professorale. Les deux exemples suivants (Merle, 2007) présentent la façon dont des comportements d'élèves particuliers peuvent orienter la dynamique relationnelle des interactions maître-élèves :

C'est vrai que tu as des élèves qui sont épouvantables aussi, ça arrive : j'avais une classe, il y avait deux mecs, un jour à l'inter-cours, j'étais seule avec eux parce que les autres étaient sortis, et il y en a un qui a dit en rigolant à son copain : « on se la coince ». Des mecs comme ça, aucune pitié. Je crois qu'il n'a pas mesuré la distance à laquelle je tenais. Pour lui dans sa tête, ça n'était même pas insultant, parce que c'est comme ça qu'il doit traiter les filles par ailleurs, mais je m'en fous, je n'ai pas non plus une pitié infinie pour ces pauvres petits enfants d'ouvriers (professeur femme, classe terminale). (pp. 86-87)

J'ai de bonnes élèves, j'ai trois petites gamines qui ont vraiment une tête d'ange, qui sont vraiment..., qui sont adorables, qui sont bosseuses, qui répondent, qui connaissent plein de trucs. Et c'est sûr que quand j'arrive à leur copie, j'ai un préjugé positif, et je le sens. Alors quelquefois, je me dis : « oh là là ! Est-ce que tu ne l'as pas surnotée ? » (professeur femme, classe de troisième). (p. 87)

Les biais d'évaluation et plus généralement les incertitudes de la notation renvoient, dans cette perspective, non à une interprétation en termes de préjugés sociaux des professeurs, mais à une question de police scolaire qui prend la forme de sanctions et de récompenses par la notation selon les comportements adoptés en classe. En fait, l'analyse des pratiques pédagogiques et des interactions en classe montre que celles-ci peuvent relever d'une explication en termes de préjugés sociaux lorsque le professeur déclare faire « surtout attention aux bons élèves », et exprime de « la répugnance pour les parents de milieux populaires ». Mais les pratiques déclarées par les maîtres peuvent aussi s'opposer à ces préjugés lorsque le professeur indique faire « attention aux moins bons », et « rechercher le contact avec les parents de milieux populaires ».

Arrangements évaluatifs et performances scolaires

Ce serait se tromper sur la signification sociale de ces arrangements évaluatifs que de les juger négativement. Dans les situations les plus ordinaires et les plus fréquentes, ces arrangements constituent une façon de « tenir » les élèves ou, par des bonnes notes, favoriser leur mobilisation scolaire. Quelques professeurs emploient d'ailleurs le terme de « notes thérapeutiques » pour désigner cette pratique (Merle, 2007). Laxisme ? Nullement. Felouzis (1997) a montré que ces « indulgences calculées » favorisaient les progressions des élèves aussi bien en mathématiques qu'en français. Il s'agit d'une modalité spécifique des « effets d'attentes » : la bonne note et la reconnaissance parentale apportée par la réussite aux devoirs sont des

sources d'encouragement, favorisent la mobilisation scolaire, redonnent du sens au travail et aux études et finalement créent les conditions d'une amélioration des compétences scolaires. *A contrario*, une notation sévère est plutôt source de découragement scolaire et aboutit le plus souvent à freiner le rythme moyen de progression des élèves. Autrement dit, l'évaluation sommative est aussi formative, pas seulement en raison des conseils qui peuvent être apportés lors de la correction et dont on connaît l'effet bénéfique sur les acquis cognitifs des élèves, mais aussi par le niveau même de la note, plus ou moins source de mobilisation scolaire.

Cependant, si certains types d'arrangements évaluatifs semblent exercer des effets positifs sur le niveau des performances scolaires, une telle relation est variable selon l'établissement d'exercice. Ainsi, dans certains « grands lycées » des centres-villes, des moyennes généralement assez basses sont attribuées à des lycéens de niveau moyen afin d'assurer, via le redoublement ou le changement d'établissement, un taux de réussite au baccalauréat proche de 100 %. La publication des palmarès des établissements favorise la concurrence entre établissements et incite les chefs d'établissement à une vigilance accrue lors du passage en classes de première et terminale. Ce contexte scolaire particulier suscite l'élitisme et, par ricochet, des notations plus sévères au détriment des élèves faibles ou moyens. Le projet d'être un « bon établissement », au sens réducteur mais commun de l'expression, peut être contradictoire avec des arrangements évaluatifs, internes à la classe, et favorables aux progrès des élèves. Le recours au terme d'« arrangement », qui évoque clairement « l'ordre négocié » des interactionnistes, ne signifie pas que les acteurs sociaux participent de façon identique à la « définition de la situation », et ne postule pas non plus l'absence d'inégalités structurelles de statut et de pouvoir à l'intérieur de l'institution scolaire. L'usage du terme met l'accent sur la diversité des situations concrètes, sur les processus qui les constituent, ainsi que sur les dynamiques contradictoires et parfois conflictuelles qui animent l'institution.

CONCLUSION

Par rapport au modèle déterministe, le modèle interactionniste de l'évaluation, centré sur des pratiques d'arrangements, présente un triple avantage : mieux rendre compte de la diversité des situations scolaires appréhendées notamment en termes de différence de sélectivité scolaire et sociale des établissements ; rendre plus facilement intelligible la réussite scolaire en milieu populaire, trajectoires sociales peu compatibles avec les explications en termes de préjugés sociaux ou « d'école reproductrice » ; être conciliable avec la diversité des origines socioprofessionnelles des enseignants, donnée peu compatible avec des stéréotypes et représentations professorales qui seraient partagés par tous de façon identique.

Si le modèle interactionniste présente une pertinence supérieure au modèle déterministe, celui-ci n'est pas pour autant caduc. Dans les sciences sociales, mais pas uniquement d'ailleurs, un paradigme interprétatif peut disposer d'un pouvoir interprétatif plus large que ses concurrents sans pour autant rendre compte de la totalité des données empiriques. De surcroît, certaines données sont susceptibles de faire l'objet d'interprétations pertinentes par des paradigmes concurrents. Il en est ainsi, par exemple, de la surnotation des élèves d'origine aisée dont rendent compte à la fois les paradigmes déterministe et interactionniste. Faute de parvenir à une falsification du paradigme déterministe, la vigilance épistémologique propre au chercheur impose de conserver ce modèle interprétatif. Il ne faut pas non plus s'empêcher de considérer que certaines situations de classe – qui resteraient à définir – pourraient être mieux interprétées par le modèle déterministe et d'autres par une modélisation interactionniste. Ce dernier modèle pourrait être également affiné. Par construction, les interactions sont situées mais également dépendantes des « carrières » des acteurs. En France, le débat actuel sur les apprentissages sans les notes, à l'instar du modèle finlandais de l'école élémentaire, oppose les thuriféraires de la notation à ceux qui, partisans du changement, en critiquent les excès et les dérives. L'histoire scolaire et professionnelle des premiers et des seconds serait susceptible d'apporter des éléments d'interprétations, d'enrichissement, de complexification et d'autofécondation des modèles déterministe et interactionniste.

Toutefois, si les paradigmes déterministe et interactionniste sont l'objet de complémentarités possibles, ils demeurent rattachés à deux épistémologies opposées. Le modèle déterministe se rattache à l'épistémologie durkheimienne de la rupture. Bachelard (1938/1989) en a présenté une formulation classique : « L'opinion pense mal ; elle ne pense pas : elle traduit des besoins en connaissances. [...] On ne peut rien fonder sur l'opinion : il faut d'abord la détruire. Elle est le premier obstacle à surmonter » (p. 14). Menées dans le cadre de PISA, les comparaisons internationales des systèmes éducatifs sont l'illustration de cette démarche d'objectivation statistique qui se construit en dehors du sens commun, parfois même en rupture avec celui-ci (voir la contribution de Lafontaine & Monseur dans le présent ouvrage).

A contrario, les approches interactionnistes, fondées le plus souvent sur des entretiens, des analyses de type clinique ou des observations ethnographiques, adhèrent pour une grande part à une épistémologie continuiste. Dans cette épistémologie, il n'existe pas de rupture définitive et infranchissable entre la connaissance et l'opinion et, surtout, entre la connaissance et l'expérience des acteurs. Sans être forcément un « savant de l'intérieur », l'acteur ou le sujet détient une forme de savoir sur sa pratique. Il dispose d'une expertise de type professionnel et sollicite plus ou moins implicitement une épistémologie pratique. Dans les approches interactionnistes, la connaissance peut éventuellement se construire contre l'expérience mais elle en est le plus souvent une émanation retravaillée,

généralisée et typologisée. Un exemple permet d'illustrer ce positionnement épistémologique qui postule l'existence d'un continuum entre le sens savant et le sens commun. Alors qu'élèves et parents ont souvent attribué l'échec ou la réussite en classe à la qualité du maître, la sociologie de l'école a longtemps ignoré cette variable avant de découvrir « l'effet maître ». Celui-ci n'est rien d'autre que la traduction scientifique et la redécouverte de l'opinion commune du « bon » et du « mauvais maître » (Merle, 2004). Dans la tradition interactionniste à laquelle se rattache notamment l'analyse des négociations maître-élèves dans la classe, la connaissance des pratiques professorales, en l'occurrence d'évaluation, se construit ainsi en collaboration, plus ou moins distanciée, avec les maîtres, leur histoire, leur culture, leur identité personnelle, professionnelle, voire disciplinaire. En ce sens, si les modèles déterministe et interactionniste participent de façon complémentaire à une science de l'évaluation, celle-ci n'a pas de fondement épistémologique unifié. Cette diversité, loin d'être un obstacle à la connaissance ou à la constitution d'une « science de l'évaluation », est un gage de richesses dans l'investigation et les résultats.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Ansart, P. (1990). *Les sociologies contemporaines*. Paris : Seuil.
- Aymes, J. (1979). Une expérience de multi-correction. *Bulletin de l'association des professeurs de mathématiques de l'enseignement public*, 321, 789-797.
- Bachelard, G. (1938/1989). *La formation de l'esprit scientifique*. Paris : Vrin.
- Barrère, A. (2002). *Les enseignants au travail*. Paris : Presses Universitaires de France.
- Boudon, R. (1984). *La place du désordre : critiques des théories du changement social*. Paris : Presses Universitaires de France.
- Chevallard, Y. (1991). Vers une analyse didactique des faits d'évaluation. In J.-M. De Ketele (Éd.), *L'évaluation : approche descriptive ou prescriptive ?* (pp. 31-59). Bruxelles : De Boeck.
- De Landsheere, G. (1976). *Évaluation continue et examens. Précis de didactologie*. Paris : Labor.
- Durkheim, E. (1893/1994). *De la division du travail social*. Paris : Presses Universitaires de France.
- Durkheim, E. (1895/1986). *Les règles de la méthode sociologique*. Paris : Presses Universitaires de France.
- Duru-Bellat, M. (1995). Filles et garçons à l'école, approches sociologiques et psycho-sociales. *Revue française de pédagogie*, 110, 75-109.
- Duru-Bellat, M. & Mingat, A. (1993). *Pour une approche analytique du fonctionnement du système éducatif*. Paris : Presses Universitaires de France.
- Felouzis, G. (1994). *Le collègue au quotidien*. Paris : Presses Universitaires de France.

- Felouzis, G. (1997). *L'efficacité des enseignants. Sociologie de la relation pédagogique*. Paris : Presses Universitaires de France.
- Laugier, H. & Weinberg, D. (1936). *Commission française pour l'enquête Carnegie sur les examens et concours. La correction des épreuves écrites au baccalauréat*. Paris : Maison du livre.
- Merle, P. (2004). Mobilisation et découragement scolaires. L'expérience subjective des élèves. *Éducation et société. Revue internationale de sociologie*, 13, 193-208.
- Merle, P. (2007). *Les notes. Secrets de fabrication*. Paris : Presses Universitaires de France.
- Noizet, G. & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Oget, D. (1999). *Efficacité et coûts du baccalauréat général et technologique : quelle alternative à l'organisation des épreuves ?* Thèse de doctorat, Université de Bourgogne.
- Pourtois, J.-P. (1978). Le niveau d'expectation de l'examineur est-il influencé par l'appartenance sociale de l'enfant ? *Revue française de pédagogie*, 44, 34-37.
- Rosenthal, R. & Jacobson, L. (1971). *Pygmalion à l'école*. Paris : Casterman.
- Sirota, R. (1988). *L'école primaire au quotidien*. Paris : Presses Universitaires de France.
- Suchaut, B. (2008). La loterie des notes au bac. Un réexamen de l'arbitraire de la notation des élèves. *Les Documents de Travail de l'IREDU*. Consulté le 15 mai 2011 dans <http://hal.inria.fr/docs/00/26/09/58/PDF/08005.pdf>
- Weber, M. (1921/1995). *Économie et société*. Paris : Plon.
- Weber, M. (1964). *L'éthique protestante et l'esprit du capitalisme*. Paris : Plon.

Élaboration et validation d'un questionnaire d'attitudes en éducation : considérations épistémologiques et métrologiques

**Pierre Valois,
Claude Houssemand & Astrid de Leeuw
Université du Luxembourg**

INTRODUCTION

Un des problèmes majeurs auxquels sont confrontés les chercheurs et les professionnels des sciences de l'éducation en général tient à la difficulté de mesurer de façon adéquate les caractéristiques des individus et des groupes qui les intéressent. Quelle que soit la propriété visée (attitude envers l'école, compétence dans une discipline donnée, habileté, etc.), il leur faut, en effet, imaginer des façons de faire qui leur permettent finalement de classer les individus selon les degrés ou niveaux auxquels ces derniers possèdent cet attribut ou cette qualité particulière. Cette opération n'est pas évidente en soi. En dépit de son apparente familiarité, elle doit, en effet, s'inscrire dans une démarche objective, éloignée de l'intuition. La difficulté de la mesure n'est pourtant pas le seul lot de ceux et celles qui travaillent ou œuvrent dans les sciences de l'éducation. Tout récemment, l'Organisation européenne pour la recherche nucléaire de Genève a dévoilé les résultats d'une expérience visant à démontrer l'existence du boson de Higgs, une particule élémentaire qui permettrait d'expliquer pourquoi toutes les particules, comme les protons ou les électrons, ont une masse (Larousserie, 2011). Le problème de ces chercheurs consiste en ce qu'ils ne peuvent observer ce boson de manière directe. Ils doivent donc trouver des indicateurs de sa présence en repérant sa signature à travers le produit de sa désintégration. Cette situation s'apparente en quelque sorte à celle vécue par un enseignant qui doit trouver des indicateurs

ou des traces de la compétence de ses élèves à maîtriser, par exemple, la langue française. Il existe cependant une différence notable entre les deux cas, puisque contrairement au boson de Higgs qui pourrait éventuellement être directement observé, la compétence à lire en français ne le sera jamais.

La difficulté de la mesure est davantage ressentie et ubiquitaire dans certains domaines particuliers du savoir comme les sciences de l'éducation (Voyer, 2006) et cette difficulté tiendrait, pour une part du moins, au fait que contrairement à certaines propriétés physiques, les caractéristiques qui intéressent les chercheurs en éducation n'ont pas d'existence propre, du moins selon la conception positiviste. On les étiquette d'ailleurs souvent de « construits » pour bien souligner le fait qu'elles n'existent que par définition. La compétence à communiquer de façon appropriée, l'habileté en calcul, le civisme, l'attitude envers son enseignant n'existent pas en soi, mais seulement au moment où l'on accepte de les définir.

Le concept de mesure peut être vu ou analysé à travers deux grands courants épistémologiques, le courant classique (Martin, 2003) ou déterministe (Mari, 1997) et le courant représentationnel. Dans le premier, l'acte de mesurer consiste à identifier des valeurs numériques inscrites dans la nature des choses. Cette vision épistémologique de la mesure est basée sur le postulat que « toute chose accessible par l'entremise de la connaissance possède un nombre, puisque sans nombre nous ne pouvons rien comprendre ni savoir » (Mari, 1997, p. 79 ; traduit d'un extrait de l'école pythagoricienne d'environ 350 ans av. J.-C.). Une mesure correspond à l'action d'évaluer une grandeur par comparaison à une grandeur de même espèce servant d'étalon ou prise comme unité de référence. Mesurer consisterait donc essentiellement à comparer une grandeur inconnue à une connue de même nature. À n'en point douter, cette très pure définition de la mesure convient seulement pour quelques propriétés relativement simples et physiques d'objets concrets. C'est ainsi que nous pouvons mesurer la longueur d'une table à l'aide d'une autre longueur de référence (le centimètre ou le pouce). Lorsque nous y regardons de plus près cependant, nous devons bien admettre que, lorsqu'il est question de mesure dans la réalité de tous les jours, la référence à une grandeur de même nature constitue davantage l'exception que la norme et que, même en sciences physiques, les occasions sont rares où les procédés de mesure s'appuient véritablement sur une unité de référence qui soit de même nature que la propriété mesurée.

Cette conception ontologique du score vrai déterminé par une mesure a donc été contestée par le philosophe et mathématicien Russell en 1903 et plus tard par le physicien anglais Campbell en 1920. Ce dernier a proposé une vision de la mesure selon laquelle « mesurer c'est attribuer des numéros pour représenter des propriétés » (p. 267). Toutefois, pour éviter que cette opération ne soit qu'une simple numérotation plutôt qu'une véritable mesure, il ajoute la contrainte voulant que les grandeurs étudiées soient additives et que cela soit vérifiable expérimentalement. Un peu plus

tard, Stevens (1946) édulcore cette vision de la mesure en abandonnant ces exigences supplémentaires. Sur la base d'une position operationaliste, il propose alors la définition générale suivante : « Au sens large la mesure est définie comme l'attribution de nombres aux objets ou aux événements suivant des règles » (p. 677). Comme le dit Martin (2003), cette définition proposée par Stevens a fait en sorte que maintenant « les sciences de la nature (physique essentiellement) ne sont pas seules à pouvoir se prévaloir de la possibilité de mesurer, quantifier, leurs objets » (p. 507).

En dépit des nombreux écrits sur le dualisme épistémologique du concept de mesure (par exemple Mari, 2000 ; Mari, 2003 ; Michell, 2003 ; Rossi, 2007), aucune convergence absolue entre les chercheurs n'a émergé à ce jour. Comme le soulignent Laurencelle et Ramsay (2001), certains chercheurs actuels estiment que tout se mesure alors que d'autres souscrivent plutôt à l'idée que les mesures sont illusoire et que l'interprétation et les opérations faites sur celles-ci sont infondées et abusives ; quelque part entre ces deux points de vue se situent les chercheurs qui pensent que seuls certains traits peuvent se mesurer. Malheureusement, la science ne peut pas et ne pourra jamais permettre aux chercheurs de trancher avec une absolue certitude ce débat incontournable mais dans une certaine mesure insoluble. De façon plus pratique, Laurencelle et Ramsay posent la question fondamentale suivante : « Lorsqu'on administre à quelqu'un un test, un examen, une épreuve et qu'on en obtient le score X, ce résultat numérique a-t-il un sens, et y a-t-il quelque chose, une "grandeur" quelconque, dont on puisse dire qu'elle a été mesurée par cette opération ? » (p. 42).

La réponse à cette question est affirmative à la condition d'endosser certaines positions épistémologiques. Il faut d'abord accepter que la notion de mesure retenue dans les sciences humaines et les sciences sociales est plus large ou plus souple que celles utilisées dans les sciences exactes. C'est ainsi que certains chercheurs acceptent aujourd'hui de considérer comme une mesure des opérations qui, non seulement laissent tomber l'exigence d'une référence à une unité formelle de comparaison de même nature, mais encore qui n'impliquent même pas l'usage d'une unité formelle de comparaison (Finkelstein, 2009). À la suite de la théorie axiomatisée de la mesure de Suppes et Zinnes (1963), « la mesure n'est plus un préalable à la théorisation » (Martin, 2003, p. 508). Autrement dit, la mesure fait partie intégrante de la théorie, elle est une construction d'un modèle numérique qui peut être validé par des critères externes.

En clair, puisque les construits ne sont pas directement observables en sciences humaines et sociales, il nous faut donc imaginer une stratégie pour leur donner vie, pour les rendre manifestes afin de pouvoir procéder à des observations. En ce sens, Mari (2003) se demande si la science de la mesure ne consisterait pas en elle-même en un exercice méthodologique puisque : « Les mesures sont le résultat d'opérations que l'on considère adéquates pour obtenir les informations recherchées au sujet de caractéristiques

déterminées des objets de mesure » (p. 24). Dans la même veine, Finkelstein (2009) affirme que les concepts de fidélité, validité et généralisabilité sont centraux à la notion élargie de mesure (*widely-defined measurement*) et que la mesure doit être accompagnée d'une mise à l'épreuve de modèles explicatifs et prédictifs (*nomological networks*, selon Cronbach & Meehl, 1955), pour lui donner un sens comme le suggèrent Laurencelle et Ramsay (2001). À cet égard, le comportement adéquat des données psychométriques dans plusieurs modèles prédictifs fournit des arguments en faveur de la thèse voulant qu'il soit possible de mesurer en sciences sociales et humaines.

Dans les paragraphes qui suivent, nous allons tenter de cerner le concept de mesure, d'en nuancer la portée selon les circonstances et la nature des caractéristiques étudiées et de décrire brièvement les étapes principales qui caractérisent le développement d'un outil de mesure. De la problématique d'une recherche à la présentation de ses conclusions, différentes étapes obligent le chercheur à faire des choix. Cinq grandes étapes du processus de modélisation de l'instrument de mesure peuvent être différenciées, et la valeur de l'instrument dans son ensemble dépendra de la qualité des choix et des décisions effectués lors de ce processus d'élaboration et de mise au point (Voyer, 1996). De façon spécifique, les cinq étapes générales sont les suivantes :

1. La détermination très précise de ce que l'on veut mesurer et à quelle fin ;
2. La fixation du format général de l'instrument de mesure ;
3. La génération d'une banque initiale d'items ;
4. La vérification de la fidélité de l'instrument de mesure auprès d'un échantillon limité de sujets ;
5. La vérification de la validité de l'échelle.

Il est important de noter qu'il n'existe pas de suites bien définies d'opérations qui garantissent le succès ou le développement d'un instrument de mesure. Aussi est-il utile de garder un regard critique sur les choix opérationnels effectués et d'être conscient des apports, mais aussi des potentielles limites, des stratégies mises en œuvre.

Une recherche réalisée dans des écoles primaires au Québec servira d'exemple pour illustrer certaines considérations et certains jugements essentiels effectués tout au long de ce processus de modélisation. Le concept central de cette étude était celui de l'attitude envers la bonne alimentation, tel qu'il est défini dans la théorie du comportement planifié d'Ajzen (2005 ; voir aussi Fishbein & Ajzen, 2010). La théorie ne nous aide pas seulement à circonscrire notre concept ou à assurer une certaine portée scientifique à nos efforts, elle permet encore d'identifier un réseau de relations, nommé réseau nomologique, à partir duquel nous pourrions porter un certain jugement sur la valeur de notre instrument. Selon Cronbach et Meehl (1955), ce réseau inclut le cadre théorique de ce que le chercheur essaie de mesurer, un cadre empirique qui précise comment il va le mesurer, de même que la spécification

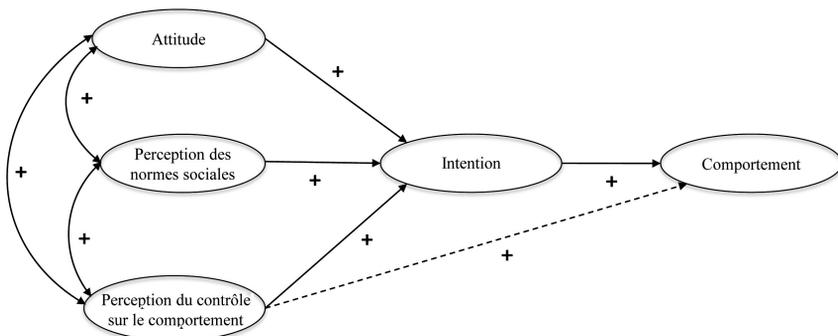
des liens au sein de et entre ces deux cadres. Dans l'une de nos analyses, nous allons vérifier, à l'aide d'analyses par équations structurales, la vraisemblance du réseau nomologique auquel appartiendrait un concept, celui d'attitude.

En résumé, à l'aide de l'exemple de l'attitude envers la bonne alimentation, nous illustrons le développement d'un instrument et présentons quelques réflexions épistémologiques qui s'y rattachent.

CONSTRUIT À MESURER ET FORMAT GÉNÉRAL DE L'OUTIL DE MESURE

La théorie du comportement planifié repose sur cinq concepts : le comportement, l'intention d'adopter le comportement, l'attitude envers l'adoption du comportement, la perception des normes sociales et la perception du contrôle sur le comportement. La figure 1 permet de visualiser le type de relations censées exister entre les concepts du modèle. Au départ de cette chaîne, nous avons l'attitude, la perception des normes sociales et la perception du contrôle sur le comportement (PCC). Bien qu'elles puissent être corrélées entre elles, ces trois variables sont supposées être conceptuellement indépendantes. Toutefois, leurs actions combinées déterminent l'intention d'une personne à adopter un comportement donné. L'intention, une fois formée, se traduit en action, variable ultime à prédire. L'intention joue ainsi le rôle de variable médiatrice entre les trois concepts de départ et le comportement. Par le biais de cette variable médiatrice, l'attitude, la perception des normes sociales et la PCC affectent indirectement le comportement. Selon Ajzen (1991), « plus les gens croient qu'ils disposent de ressources et d'opportunités et moins ils anticipent d'obstacles et empêchements plus ils perçoivent avoir du contrôle sur le comportement » (p. 196). De plus, lorsque le contrôle perçu et le contrôle effectif ou réel coïncident, la PCC peut alors exercer un effet direct sur le comportement (la flèche en pointillé de la figure 1).

Figure 1 : Illustration de la théorie du comportement planifié



L'attitude d'une personne à l'égard d'un comportement (par exemple avoir une alimentation équilibrée) est déterminée par ses croyances concernant les conséquences de l'adoption du comportement (par exemple, avoir une alimentation équilibrée m'évitera de prendre du poids) et l'évaluation de l'importance accordée à chacune des conséquences identifiées (par exemple, c'est très important pour moi de ne pas prendre de poids). Cette définition s'exprime à l'aide de l'équation suivante :

$$A = \sum_{i=1}^n c_i \times e_i$$

où A = l'attitude de la personne à l'égard du comportement ;

c_i = la croyance de la personne concernant les chances que l'adoption du comportement provoquera chacune des conséquences associées à l'adoption du comportement en cause ;

e_i = l'évaluation positive ou négative de chacune des conséquences par la personne ;

n = le nombre de conséquences.

Un exemple d'opérationnalisation de ce modèle d'attitude est présenté ci-après. La grille de réponse est conçue de sorte à faciliter la tâche des élèves de 1^{re} année primaire. Le format de la grille suggère à l'élève qu'il doit d'abord réfléchir globalement sur son accord ou désaccord avec l'énoncé et ensuite affiner cette réponse.

Croyance : Au cours des prochains mois, si je mangeais des aliments appartenant aux quatre groupes alimentaires, cela me permettrait d'être en bonne santé.

EN DÉSAACCORD			D'ACCORD		
Fortement -3	Moyennement -2	Légèrement -1	Légèrement +1	Moyennement +2	Fortement +3

Valeur : Pour moi, c'est important de manger de façon équilibrée pour être en bonne santé.

EN DÉSAACCORD			D'ACCORD		
Fortement -3	Moyennement -2	Légèrement -1	Légèrement +1	Moyennement +2	Fortement +3

Pour comprendre les problèmes liés à ce type de mesure, le lecteur pourra se reporter à Valiquette, Valois, Desharnais et Godin (1988).

Le construit et la fixation du format général de l'échelle d'attitude étant définis, il faut maintenant passer aux étapes subséquentes, soit a) la génération d'une banque initiale d'items ; b) la vérification de la fidélité de l'instrument de mesure auprès d'un échantillon limité de sujets ; et c) la vérification de la validité de l'échelle. Trois études correspondant respectivement à chacune de ces trois étapes sont décrites ci-après.

GÉNÉRATION DES ITEMS : ÉTUDE QUALITATIVE

Une étude qualitative a été réalisée pour sélectionner un premier ensemble d'items composant le questionnaire d'attitudes (Ajzen & Fishbein, 1980). Cette étape est primordiale car elle permet d'identifier les croyances réelles de la population visée par l'étude et ainsi éviter que le questionnaire soit biaisé. Par exemple, il pourrait arriver qu'un chercheur sélectionne ses items à partir de questionnaires validés, mais que les validations aient été réalisées auprès d'échantillons non représentatifs de la population qu'il étudie. Dans un tel cas, il y a lieu de penser que les croyances retenues ne reflètent pas, du moins en totalité, les croyances dominantes de la population-cible. Il pourrait aussi arriver que le questionnaire soit biaisé parce que le chercheur a lui-même formulé les croyances, sur la base de son expérience personnelle ou de ses intuitions théoriques. Encore ici, il est fort probable que ces croyances ne concordent pas avec les croyances dominantes de la population-cible. En résumé, cette étape qualitative est primordiale dans le développement d'un questionnaire d'attitudes et trop souvent omise par les chercheurs qui ont recours à ce type d'instrument de mesure.

Échantillon

Afin d'identifier les croyances dominantes des enfants du primaire quant à l'adoption d'une bonne alimentation, 12 élèves (6 filles et 6 garçons), ayant des caractéristiques similaires à celles de la population ciblée par l'étude ont été recrutés dans une classe de 1^{re} année primaire avec la permission des parents, du directeur d'école et de l'enseignante. Ils étaient tous âgés entre 10 et 11 ans.

Instruments de mesure et procédure de collecte des données

La passation du questionnaire a eu lieu dans le cadre d'un cours pour lequel l'enseignante avait accordé 30 minutes aux chercheurs. Après avoir expliqué l'objectif de l'étude aux élèves, il leur a été demandé de répondre, par écrit,

à deux questions permettant d'identifier leurs perceptions des avantages et désavantages associés à l'adoption d'une bonne alimentation : 1) Certaines personnes s'alimentent bien. Selon toi, quel intérêt ou quel plaisir ces personnes trouvent-elles dans le fait de bien s'alimenter ? ; 2) Certaines personnes s'alimentent mal. Selon toi, quel intérêt ou quel plaisir ces personnes trouvent-elles dans le fait de mal s'alimenter ?

Résultats

L'analyse de contenu de cette information a été réalisée par une éducatrice en nutrition expérimentée dans la construction et la validation de questionnaires. Tel que suggéré par Ajzen (1985), les croyances les plus fréquemment mentionnées par les élèves ont été retenues et incluses dans la première version du questionnaire d'attitudes. L'analyse des réponses a permis d'identifier sept croyances dominantes des jeunes à l'égard des conséquences associées à une bonne alimentation. Plus précisément, ceux-ci croient qu'une bonne alimentation permet : a) d'être en bonne santé ; b) d'être en pleine forme ; c) d'avoir de belles dents ; d) d'avoir moins de maladies ; e) de ne pas engraisser ; f) de mieux grandir ; g) d'avoir une alimentation qui a bon goût. C'est à partir de ces croyances dominantes que le questionnaire d'attitudes construit selon la théorie du comportement planifié a été élaboré et que ses qualités métrologiques ont été éprouvées.

Considérations épistémologiques

Dans cette première étude, les réponses des élèves aux questions ouvertes ont été analysées par une seule personne, ce qui n'est pas souhaitable, du moins sur la base des théories de la mesure. En effet, hormis la confiance qu'un chercheur pourrait avoir envers la personne qui effectue l'analyse de contenu en raison de la qualité de ses évaluations lors d'une formation d'entraînement à cette méthode, rien ne nous indique alors que celle-ci a bien saisi le sens des réponses des élèves ou que son interprétation n'est pas biaisée. Il est donc souhaitable d'avoir recours à plus d'un expert et d'estimer la fidélité inter-juges à l'aide de différentes méthodes d'analyse (coefficient de corrélation des rangs de Spearman [ρ] et le coefficient de corrélation intra-classe [CII], par exemple).

En somme, il apparaît qu'une combinaison d'études qualitative et quantitative est souhaitable lors de la construction d'un questionnaire d'attitudes. En effet, alors que l'étude qualitative permet d'identifier les croyances dominantes réelles de la population-cible, l'étude quantitative permet de vérifier la qualité de l'analyse de contenu des réponses aux questions ouvertes effectuée par les experts. À notre avis, l'opposition entre « positivisme » et « constructivisme » crée une certaine confusion dans

le débat épistémologique sur le concept de mesure. Les deux approches sont plutôt complémentaires qu'opposées en ce sens que l'harmonisation de leurs forces respectives favorise l'objectivité de la mesure. En effet, l'approche qualitative permet d'éliminer les biais que le chercheur pourrait introduire, volontairement ou non, en choisissant lui-même les items composant l'outil de mesure. Dans la présente étude, en ayant eu recours à des questions ouvertes, nous avons fait en sorte que les participants rapportent leurs perceptions des avantages et désavantages associés à une alimentation équilibrée, sans être confinés à un cadre de réponse prédéfini. En ce faisant, nous nous sommes ainsi assurés de faire ressortir l'opinion réelle des jeunes, leurs croyances dominantes par rapport à l'alimentation équilibrée. Par ailleurs, l'étude quantitative qui suit a permis de vérifier, à l'aide d'indices métrologiques, si le questionnaire d'attitudes composé des croyances dominantes identifiées à cette étape présentait un bon indice de fidélité. Ces résultats sont présentés dans la section suivante.

PRÉTEST DU QUESTIONNAIRE : ESTIMATION DE LA FIDÉLITÉ DE L'ÉCHELLE D'ATTITUDE

L'objectif de ce prétest du questionnaire est double : a) sélectionner les items dont le contenu est clair et pertinent ; b) retenir les items qui portent sur l'essence même de l'objet d'attitude mesuré. En d'autres mots, le questionnaire d'attitudes doit faire preuve d'un haut degré de cohérence interne (par exemple un coefficient d'alpha de Cronbach élevé).

Échantillon

Le questionnaire a été administré à 34 élèves âgés entre 9 et 12 ans ($M = 10.25$, $E.-T. = 0.54$), dont 18 filles et 16 garçons. Ces élèves ont été recrutés dans 10 écoles primaires au Québec offrant un enseignement de troisième cycle primaire et n'étant pas impliquées dans d'autres projets de recherche au moment de l'étude.

Instruments de mesure et procédure de collecte des données

Les élèves ont répondu anonymement au questionnaire d'attitudes envers l'alimentation en classe. Un accord écrit d'au moins un des parents était nécessaire pour que l'élève puisse participer à l'étude. Le questionnaire d'attitudes comprenait sept paires d'items et les élèves devaient indiquer sur une échelle de réponse en six points, selon le modèle présenté auparavant, leur degré d'accord avec les sept croyances suivantes : Au cours des prochains mois, si je mangeais des aliments appartenant aux quatre groupes

alimentaires et selon les portions recommandées par *Le Guide alimentaire canadien*¹, cela me permettrait : a) d'être en bonne santé ; b) d'être en pleine forme ; c) d'avoir de belles dents ; d) d'avoir moins de maladies ; e) de ne pas engraisser ; f) de mieux grandir ; g) d'avoir une alimentation qui a bon goût. Ils devaient ensuite indiquer leur degré d'accord sur la même échelle de réponse (- 3, totalement en désaccord à +3, totalement d'accord) avec les sept valeurs correspondant aux sept croyances, respectivement : Pour moi, c'est important de bien m'alimenter : a) pour être en santé ; b) pour être forme, etc.

Résultats

Les résultats indiquent que les élèves ont une attitude assez favorable envers le comportement qui consiste à s'alimenter de façon équilibrée ($M = 5.91$ sur un maximum de 9, $E.-T. = 2.73$). Les résultats révèlent aussi que le questionnaire d'attitudes possède une bonne fidélité, plus précisément une bonne consistance interne (alpha de Cronbach = .80) et que tous les items semblent mesurer le même concept, les indices de discrimination des sept combinaisons d'items variant de .36 à .63. Toutefois, les items 2 et 4 ont été retirés de l'échelle d'attitude puisque leur contenu était redondant avec celui de l'item 1, ce qui provoque une surestimation de la fidélité de l'échelle. Après le retrait de ces deux items, le coefficient alpha est descendu à .72, ce qui est très acceptable. Le questionnaire d'attitudes utilisé dans l'étude de validation suivante comprend donc cinq des sept items de départ.

Considérations épistémologiques

En dépit de ce que les auteurs écrivent souvent, une échelle d'attitude dont le coefficient de consistance interne oscille autour de .80 représente l'exception plutôt que la norme (Voyer, 1996). En effet, très souvent les échelles d'attitude utilisées dans les recherches présentent des coefficients autour de .70. À nos yeux cependant, cette dernière valeur constitue un minimum dont il ne faut pas trop abuser. Cette dernière remarque est bien sûr fonction du nombre d'items composant l'échelle d'attitude, car l'obtention d'un coefficient alpha de .90 pour une échelle d'attitude ne comprenant que trois items serait, sans aucun doute, le signe d'une redondance des items.

Néanmoins, il reste important de considérer qu'il sera toujours préférable d'avoir une échelle dont le coefficient de consistance interne est légèrement plus faible, mais pour laquelle il existe une certaine assurance quant à la validité, que l'inverse. Il est ainsi essentiel de s'assurer que la variable

1. Santé et Bien-être social Canada (1992a, 1992b).

mesurée par notre instrument correspond bien à la variable latente visée (c'est-à-dire l'attitude envers la bonne alimentation dans la présente étude). Il faut donc se préoccuper de sa validité, comme l'expliquent Laveault et Grégoire (2008) : « La précision de la mesure est certes importante, mais elle est inutile si le test n'évalue pas, ou évalue mal, la réalité visée par ses concepteurs » (p. 163).

ÉTUDE QUANTITATIVE : VALIDITÉ DE L'ÉCHELLE D'ATTITUDE

Le but principal de cette étude consiste à vérifier la validité de notre construit à l'aide d'équations structurales. La modélisation par équations structurales (MES), parfois appelée analyse en pistes causales, peut aider le chercheur à mieux connaître les liens causaux entre les variables. L'un des avantages de la MES est sa capacité de prendre en compte l'erreur de mesure dans le processus d'estimation de plusieurs relations dépendantes.

Dans le cadre du processus de validation du questionnaire d'attitudes, les analyses viseront à éprouver sa validité nomothétique. Il s'agit en fait d'évaluer le réseau des relations entre le concept d'attitude et d'autres variables, tel que présenté dans la figure 1. À l'aide du logiciel *Mplus* 5.2 (Muthén & Muthén, 1998/2007), l'analyse par équations structurales a été réalisée. Pour procéder à ces analyses, il est recommandé d'utiliser plusieurs indicateurs pour chaque variable, car les scores de plusieurs indicateurs tendent à être plus fiables et valides que ceux d'un seul indicateur (Kline, 2011). Ici, toutes les variables, sauf le comportement, ont deux indicateurs représentant le score aux items pairs et impairs de l'échelle (Bandalos & Finney, 2001 ; Nasser & Takahashi, 2003). Le comportement se manifeste à travers trois indicateurs correspondant à trois mesures de comportement alimentaire à trois moments différents.

Pour apprécier l'ajustement du modèle aux données, différents indices ont été utilisés : le ratio S-B χ^2/dl , l'indice d'ajustement comparatif (CFI), l'indice Tucker-Lewis (TLI), ainsi que la racine carrée de l'approximation de l'erreur type (RMSEA). Selon les suggestions de Kline (2011), un bon ajustement du modèle aux observations empiriques peut être reporté si le ratio S-B χ^2/dl est inférieur à 3, si les indices CFI et TLI ont une valeur égale ou supérieure à .90, et si la valeur RMSEA est inférieure à .05 (bon ajustement) ou à .08 (ajustement acceptable).

Échantillon

La vérification de la validité du questionnaire d'attitudes par l'entremise de ses liens avec les autres variables de la théorie du comportement planifié a

été réalisée à l'aide d'un échantillon de 275 élèves âgés entre 9 et 12 ans ($M = 10.33$, $E.T. = 0.54$). Parmi eux (148 filles et 127 garçons), 84.7 % connaissaient *Le Guide alimentaire canadien pour manger sainement* (Santé et Bien-être social Canada, 1992a).

Instruments de mesure et procédure de collecte des données

Les élèves ont été recrutés dans 10 écoles primaires et la passation du questionnaire s'est effectuée dans les mêmes conditions que lors du pré-test. Toutes les variables de la théorie du comportement planifié ont été mesurées au regard de l'adoption d'un comportement alimentaire équilibré. Toutes ces variables ont été mesurées au début du semestre scolaire, sauf le comportement dont des mesures ont aussi été réalisées quatre et sept mois plus tard. Cette décision permettait d'atténuer les erreurs de mesure du comportement alimentaire des élèves.

Comportement : Les élèves devaient rapporter ce qu'ils avaient mangé dans la journée précédant les trois passations du questionnaire. Leurs réponses ont été analysées par une nutritionniste. Le score varie entre 1 et 4, ces scores correspondant au nombre de groupes alimentaires différents consommés au cours d'une journée.

Intention : Elle désigne jusqu'à quel point l'élève est déterminé à manger de façon équilibrée. Elle est mesurée au moyen de sept questions. Par exemple, une question consistait à demander aux élèves d'indiquer sur une échelle de type Likert variant de 1 (totalement en désaccord) à 6 (totalement d'accord) s'ils avaient l'intention de manger des aliments d'au moins trois des quatre groupes du *Guide alimentaire canadien* (*op. cit.*) à chaque repas. L'alpha de Cronbach de cette échelle est de .80.

Attitude : Pour la mesure de l'attitude, le questionnaire à cinq items décrit dans l'étude précédente a été utilisé.

Perception des normes sociales : Les élèves devaient d'abord exprimer leurs croyances par rapport aux attentes de six personnes-références (par exemple les amis, les parents) concernant l'adoption d'un comportement alimentaire équilibré. Pour cela, ils devaient indiquer sur une échelle de type Likert semblable, s'ils pensaient que chacune de ces personnes s'attendait à ce qu'il mange chaque jour de façon équilibrée. Ils devaient ensuite indiquer, sur une échelle de type Likert variant de fortement en désaccord à fortement d'accord, leur degré de motivation à se conformer aux attentes de chacune des personnes-références. Par la suite, chaque score de croyance a été multiplié par la motivation de l'élève à se conformer aux attentes de la personne-référence visée, la moyenne des produits donnant le score individuel des normes sociales. Le coefficient alpha de Cronbach de cette échelle est de .84.

Perception du contrôle sur le comportement : La perception du contrôle sur le comportement fait référence à la perception que l'élève a de ses capacités à surmonter les obstacles qui pourraient l'empêcher de manger de façon équilibrée. Les élèves devaient indiquer, sur une échelle de type Likert à six niveaux, leur degré d'accord ou de désaccord quant au pouvoir de quatre situations susceptibles de les décourager de s'alimenter de façon équilibrée. Par exemple, l'élève devait répondre à la question suivante : Au cours des prochains mois, même si je suis avec un groupe d'adolescents, je me sentirai capable de manger de façon équilibrée. L'échelle a démontré un degré de consistance interne acceptable, avec un alpha estimé à .60.

Résultats

Les résultats du tableau 1 indiquent que le modèle de la théorie du comportement planifié s'ajuste bien aux données.

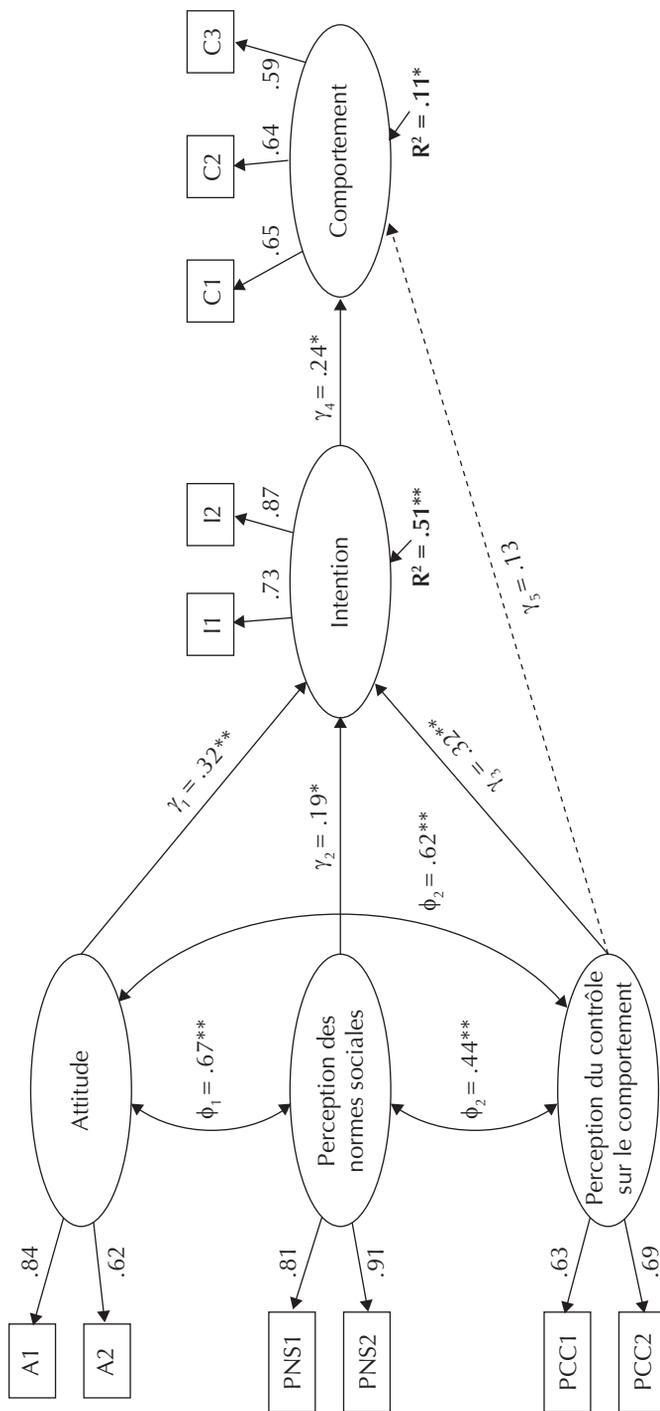
Tableau 1 : Indices d'ajustement du modèle de la théorie du comportement planifié au regard de l'alimentation équilibrée

	S-B χ^2	df	p	S-B χ^2/df	CFI	TLI	RMSEA (90 % IC)
Modèle	56.66	36	.02	1.57	.98	.97	.05 (.02 ; .07)

La figure 2 est une représentation graphique des résultats du modèle éprouvé qui indique que les liens qui vont de l'attitude (γ_1), de la perception des normes sociales (γ_2) et de la perception du contrôle comportemental (γ_3) vers l'intention sont significatifs. Ainsi, l'intention de s'alimenter de manière équilibrée est prédite par l'attitude envers la bonne alimentation ($\gamma_1 = .32, p < .01$), la perception des normes sociales ($\gamma_2 = .19, p < .05$) et la perception qu'a l'élève de son contrôle sur son alimentation ($\gamma_3 = .32, p < .01$). Selon les estimations obtenues, les trois variables contribuent à expliquer 51 % de la variance de l'intention. À ceci s'ajoute, comme stipulé par le modèle théorique, un lien direct modéré positif entre le comportement consistant à s'alimenter de manière équilibrée et l'intention de le faire ($\gamma_4 = .24, p < .05$). Or, la relation directe entre la perception du contrôle sur le comportement et le comportement lui-même est statistiquement insignifiante ($\gamma_5 = .13, ns$).

Enfin, les résultats de la figure 2 indiquent que l'attitude envers la bonne alimentation est fortement corrélée avec la perception des normes sociales ($\phi = .67, p < .01$), ainsi qu'avec la perception du contrôle sur le comportement ($\phi = .62, p < .01$), et qu'entre ces deux dernières, il y a une corrélation modérée positive ($\phi = .44, p < .01$).

Figure 2 : Résultats de l'analyse par équations structurales



Les liens observés entre le concept d'attitude et les autres variables correspondent aux postulats de la théorie du comportement planifié. À la lumière de ces résultats, il y a lieu de croire en la validité du questionnaire d'attitudes élaboré dans le cadre de cette recherche.

Considérations épistémologiques

Comme nous l'avons mentionné dans l'introduction, la mesure doit être accompagnée d'une mise à l'épreuve de modèles explicatifs et prédictifs (*nomological networks*) pour lui donner un sens. Les résultats obtenus à l'aide d'analyses par équations structurales indiquent la vraisemblance du réseau nomologique auquel appartiendrait le concept d'attitude. Toutefois, il apparaît que les logiciels d'analyses par équations structurales ne permettent pas, ou du moins rarement, de conclure qu'il y a des liens causaux entre les variables (Kline, 2005 ; Pearl, 2012).

Par contre, il est fortement recommandé de faire précéder l'étape de modélisation par équations structurales d'une analyse d'items basée sur les modèles de réponses aux items, ce que nous avons fait mais n'avons pu présenter dans le cadre de ce chapitre. L'étude de van der Mass, Molenaar, Maris, Kievit et Borsboom (2011) indique que les modèles de réponses aux items sont adéquats pour mesurer le concept d'attitude puisqu'elle fait la démonstration mathématique qu'il y a un lien entre les modèles de processus de prise de décision et les modèles de réponses aux items.

Enfin, il importe de mentionner toutefois que les modèles paramétriques de réponses aux items sont peu appliqués dans la majorité des études, compte tenu des grandes tailles d'échantillon qu'ils nécessitent. Aussi, dans le cas où les tailles d'échantillon sont peu élevées, il serait plus approprié d'utiliser les modèles non paramétriques de réponses aux items. Quelques logiciels conviviaux tels que TESTGRAF (Ramsay, 2000) et EIRT (Germain, Valois & Abdous, 2007 ; Valois, Houssemand, Germain & Abdous, 2011) permettent ce type d'analyse.

CONCLUSION

Cette étude avait pour but de présenter et discuter les étapes du processus de modélisation d'un instrument de mesure selon une notion élargie du concept de mesure, et en adoptant la position selon laquelle la mesure consiste en un exercice méthodologique. Pour ce faire, le concept d'attitude envers une alimentation équilibrée a servi d'exemple. Il se dégage de ce travail que la combinaison d'études qualitative et quantitative est souhaitable lors de la construction d'un questionnaire d'attitudes. Les présents résultats montrent que l'étude qualitative permet de s'assurer que les contenus des items

d'attitude concordent avec les croyances dominantes des personnes de la population ciblée et que les études quantitatives basées sur les modèles de mesure servent alors, d'une certaine façon, à vérifier la qualité des items proposés et la validité du questionnaire d'attitudes qui en résulte à l'aide de la théorie des tests et l'identification d'un réseau nomologique.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Ajzen, I. (1985). From intentions to actions : A theory of planned behavior. In J. Kuhl & J. Beckmann (Éd.), *Action control : From cognition to behavior* (pp. 11-39). Heidelberg : Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Ajzen, I. (2005). *Attitudes, personality, and behavior* (2^e éd.). Milton-Keynes, England : Open University Press/McGraw-Hill.
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting behavior*. Englewood Cliffs, NJ : Prentice Hall.
- Bandalos, D. L. & Finney S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Éd.), *Advanced Structural Equation Modeling : New Developments and Techniques* (pp. 269-296). Mahwah, NJ : Lawrence Erlbaum Associates.
- Campbell, N. R. (1920). *Physics : The elements*. Cambridge : Cambridge University Press.
- Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Finkelstein, L. (2009). Widely-defined measurement – An analysis of challenges. *Measurement*, 42, 1270-1277.
- Fishbein, M. & Ajzen, I. (2010). *Predicting and changing behavior : The reasoned action approach*. New York : Psychology Press.
- Germain, S., Valois, P. & Abdous, B. (2007). *EIRT : module Excel pour effectuer des estimations de la Théorie de la réponse aux items*. Consulté le 16 janvier 2012 dans <http://libirt.sourceforge.net/fr/index.html>
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2^e éd.). New York : The Guilford Press.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3^e éd.). New York : The Guilford Press.
- Larousserie, D. (2011, 14 décembre). La chasse au boson de Higgs touche, presque, à sa fin [version électronique]. *Le Monde*. Consulté le 14 décembre 2011 dans <http://www.lemonde.fr/imprimer/article/2011/12/13/1617905.html>
- Laurencelle, L. & Ramsay, J. O. (2001). À la recherche de l'« unité de mesure » en psychométrie : réflexions sur la mesure en sciences humaines. *Mesure et évaluation en éducation*, 24, 41-52.

- Laveault, D. & Grégoire, J. (2008). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2^e éd.). Bruxelles : De Boeck.
- Mari, L. (1997). The role of determination and assignment in measurement. *Measurement, 21*, 79-90.
- Mari, L. (2000). Beyond the representational viewpoint : A new formalization of measurement. *Measurement, 27*, 71-84.
- Mari, L. (2003). Epistemology of measurement. *Measurement, 34*, 17-30.
- Martin, O. (2003). Les origines philosophiques et scientifiques de la théorie représentationnelle de la mesure (1930-50). *Social Science Information, 42*, 485-513.
- Michell, J. (2003). Epistemology of measurement : The relevance of its history for quantification in the social sciences. *Social Science Information, 42*, 515-534.
- Muthén, L. K. & Muthén, B. O. (1998/2007). *Mplus User's Guide* (5^e éd.). Los Angeles : Muthén & Muthén.
- Nasser, F. & Takahashi, T. (2003). The effect of using item parcels on ad hoc goodness-of-fit indexes in confirmatory factor analysis : An example using Sarason's reactions to tests. *Applied Measurement in Education, 16*, 75-97.
- Pearl, J. (2012). *The causal foundations of structural equation modeling* (Technical Report R-370). UCLA Cognitive Systems Laboratory. Consulté le 16 janvier 2012 dans http://bayes.cs.ucla.edu/csl_papers.html
- Ramsay, J. O. (2000). *TESTGRAF Manual*. Manuscrit non publié, McGill University, Département de psychologie.
- Rossi, G. B. (2007). Measurability. *Measurement, 40*, 545-562.
- Russell, B. (1903). *Principles of mathematics*. Cambridge : Cambridge University Press.
- Santé et Bien-être social Canada. (1992a). *Le Guide alimentaire canadien pour manger sainement*. Ottawa : Approvisionnement et services Canada.
- Santé et Bien-être social Canada. (1992b). *Renseignements sur le Guide alimentaire à l'intention des éducateurs et des communicateurs*. Ottawa : Approvisionnement et services Canada.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.
- Suppes, P. & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush & E. Galanter (Éd.), *Handbook of mathematical psychology* (pp. 1-76). New York : Wiley.
- Valiquette, C. A. M., Valois, P., Desharnais, R. & Godin, G. (1988). An item-analytic investigation of the Fishbein and Ajzen multiplicative scale : The problem of a simultaneous negative evaluation of belief and outcome. *Psychological Reports, 63*, 723-728.
- Valois, P., Houssemand, C., Germain, S. & Abdous, B. (2011). An open source tool to verify the psychometric properties of an evaluation instrument. *Procedia-Social and Behavioral Sciences, 15*, 552-556.

- Van der Mass, H. L. J., Molenaar, D., Maris, G., Kievit, R. A. & Borsboom, D. (2011). Cognitive psychology meets psychometric theory : On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, 118, 339-356.
- Voyer, J.-P. (1996). *L'élaboration d'une échelle pour la mesure d'une attitude ou d'un construit psychologique*. Manuscrit non publié, Université Laval, Faculté des sciences de l'éducation.
- Voyer, J.-P. (2006). *Du concept au résultat de sa mesure*. Manuscrit non publié, Université Laval, Faculté des sciences de l'éducation.

Autorégulation et évaluation-soutien d'apprentissage

**Dany Laveault
Université d'Ottawa**

OBJECTIFS DE LA CONTRIBUTION

Au cours des 30 dernières années, le modèle de la régulation s'est avéré particulièrement fécond parmi les chercheurs francophones en évaluation fortement inspirés par les théories socio-constructivistes de l'apprentissage (Allal, 2010 ; Cardinet, 1986 ; Laveault, 2010 ; Perrenoud, 1998). Il est associé aux ajustements nécessaires pour adapter l'enseignement et l'apprentissage à la progression des élèves vers l'atteinte d'objectifs précis. Chez plusieurs chercheurs francophones, ce modèle a contribué à privilégier un alignement de l'évaluation sur la régulation des processus cognitifs autant, sinon plus que sur la régulation des activités d'apprentissage (Perrenoud, 1998).

Cette contribution a pour objectif, dans un premier temps, de présenter comment les modèles de régulation des apprentissages, et l'autorégulation en particulier, sont associés aux enjeux de l'évaluation formative chez les chercheurs francophones et à ceux de l'évaluation-soutien d'apprentissage chez les chercheurs anglo-saxons. Dans un second temps, cette contribution vise à présenter comment, au fil des résultats empiriques, un modèle d'autorégulation s'est graduellement imposé dans mes propres recherches afin de décrire et d'expliquer comment l'évaluation et l'autoévaluation peuvent soutenir la progression des élèves et développer leur capacité à exercer un meilleur contrôle sur leur apprentissage. Dans un troisième temps, il sera question de la validité théorique et concurrente du modèle développé vis-à-vis des résultats des recherches les plus récentes et de considérer la pertinence d'ouvrir les modèles d'autorégulation aux stratégies

de régulation émotionnelle qui ont fait l'objet de nombreuses études au cours des cinq dernières années.

RÔLE DE L'AUTORÉGULATION

Il existe plusieurs définitions de l'apprentissage mais, essentiellement, celles-ci se réfèrent à un processus de changement dans le potentiel intellectuel, comportemental et affectif du sujet résultant soit de l'adaptation à son milieu par l'expérience, soit de l'organisation de ses connaissances par la réflexion et la structuration. L'évaluation sert à attester de l'existence de ce changement chez l'élève et à l'induire par les moyens appropriés.

C'est en faisant prendre conscience aux intervenants du monde de l'éducation du caractère restreint et limité des apprentissages des élèves que l'évaluation s'est elle-même remise en question. D'une part, la question de la validité des résultats se pose de façon accrue lorsqu'il s'agit de bilans de compétences ou performances complexes. L'évaluation des compétences a fait ressortir l'importance d'évaluer l'intégration des connaissances, leur mobilisation dans des contextes appropriés et leur transfert à des situations nouvelles. D'autre part, se pose de façon concrète toute la problématique du développement de méthodes d'évaluation qui servent de support à l'apprentissage des compétences. À cet égard, la solution apportée par la conception traditionnelle de l'évaluation formative est rapidement apparue comme peu satisfaisante.

Du côté anglo-saxon, de nombreux efforts ont été entrepris afin de reconceptualiser l'évaluation formative (*formative assessment*) autour d'une nouvelle appellation : *évaluation-soutien d'apprentissage* (EsA ; en anglais, *assessment for learning*). Pour plusieurs auteurs anglo-saxons, les deux expressions sont interchangeables. Pour Stiggins (2005), elles présentent des différences non négligeables. Selon lui, l'EsA est une évaluation continue et non seulement une évaluation plus fréquente. Il se réfère au fait que l'évaluation joue un rôle tant dans le contexte d'une collecte ponctuelle plus ou moins instrumentée de renseignements sur l'apprentissage des élèves que dans la communication entre l'enseignant et les élèves de même qu'entre les élèves. L'EsA consiste également à informer prioritairement l'élève sur lui-même et sur sa progression au moment où celle-ci se produit. Cette conception de l'EsA met l'accent sur une évaluation qui développe le contrôle de l'élève sur ses apprentissages par rapport à une évaluation qui développerait uniquement le contrôle de l'enseignant.

Un groupe d'experts internationaux, principalement anglo-saxons, s'est entendu sur la définition suivante d'EsA : « 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 & Laveault, 2009, p. 102). Cette définition globale, tout comme les distinctions apportées par Stiggins (2005), cherche à établir un meilleur équilibre entre le rôle de l'enseignant et le rôle de l'élève. Elle propose d'étendre l'évaluation jusque dans les questionnements, dialogues et échanges continus non seulement entre l'enseignant et les élèves mais aussi entre les élèves eux-mêmes. Dans cette perspective globale, la distinction entre l'évaluation des apprentissages (*assessment of learning*), l'EsA (*assessment for learning*) et l'évaluation en tant qu'apprentissage (*assessment as learning*) tend à s'estomper tout comme la distinction entre le sommatif et le formatif. En effet, un entretien maître-élève autour d'un livret de compétences servant à préparer un plan d'enseignement individualisé peut tout aussi bien soutenir l'apprentissage de l'élève qu'un dialogue maître-élève autour d'une difficulté précise lors de la résolution d'un problème.

Malgré la multiplication des définitions de l'évaluation, il existe un accord croissant sur le fait que le soutien d'apprentissage peut s'exercer d'une multiplicité de façons : à différents niveaux et par des méthodes variées, que celles-ci soient plus ou moins formelles, plus ou moins instrumentées, plus ou moins continues et finalement plus ou moins régulées par l'élève. Les divergences apparaissent davantage dans la modulation de ces méthodes, plus particulièrement dans le rôle joué par l'élève et par l'évaluation dans le développement d'un meilleur contrôle de celui-ci sur ses apprentissages.

Cette question n'est pas sans rappeler une problématique soulevée par la distinction apportée par Scallon (2000) entre évaluation formative et évaluation formatrice, distinction apparue nécessaire pour rendre compte des résultats des travaux de Nunziati (1990). L'évaluation formatrice pose comme point de départ que les instruments privilégiés de la construction des apprentissages sont :

- L'appropriation par les élèves des critères d'évaluation des enseignants ;
- La pratique de l'autocontrôle, l'autogestion des erreurs ;
- La maîtrise des processus d'anticipation et de planification des actions.

Vial (1997, cité par Scallon, 2000) va jusqu'à remettre en question l'influence que la régulation exerce sur l'autoévaluation : « Celle-ci doit s'inscrire dans une démarche de pleine autonomie, loin de toute conformité à des normes préétablies » (p. 266). Cette position de Vial est très proche de la conception d'évaluation en tant qu'apprentissage (*assessment as learning*, Earl, 2003) qui élève l'autoévaluation au rang de compétence à développer. Pour Earl, l'évaluation en tant qu'apprentissage est au cœur de l'EsA. Il ne suffit pas de faire de l'étudiant un collaborateur du processus d'évaluation, mais un acteur critique qui fait lui-même le lien entre évaluation et apprentissage.

Le modèle de régulation permet d'unifier ces différentes définitions de l'évaluation et de représenter l'évaluation sur un continuum variant d'une régulation assumée par l'enseignant à une régulation assumée par l'élève (Laveault, 2007) dans un but qui sert principalement l'enseignant ou principalement l'élève. En réponse à la question : « Que faut-il pour que l'évaluation soutienne l'apprentissage ? » l'EsA navigue entre faire de meilleures évaluations des élèves ou faire des élèves de meilleurs évaluateurs.

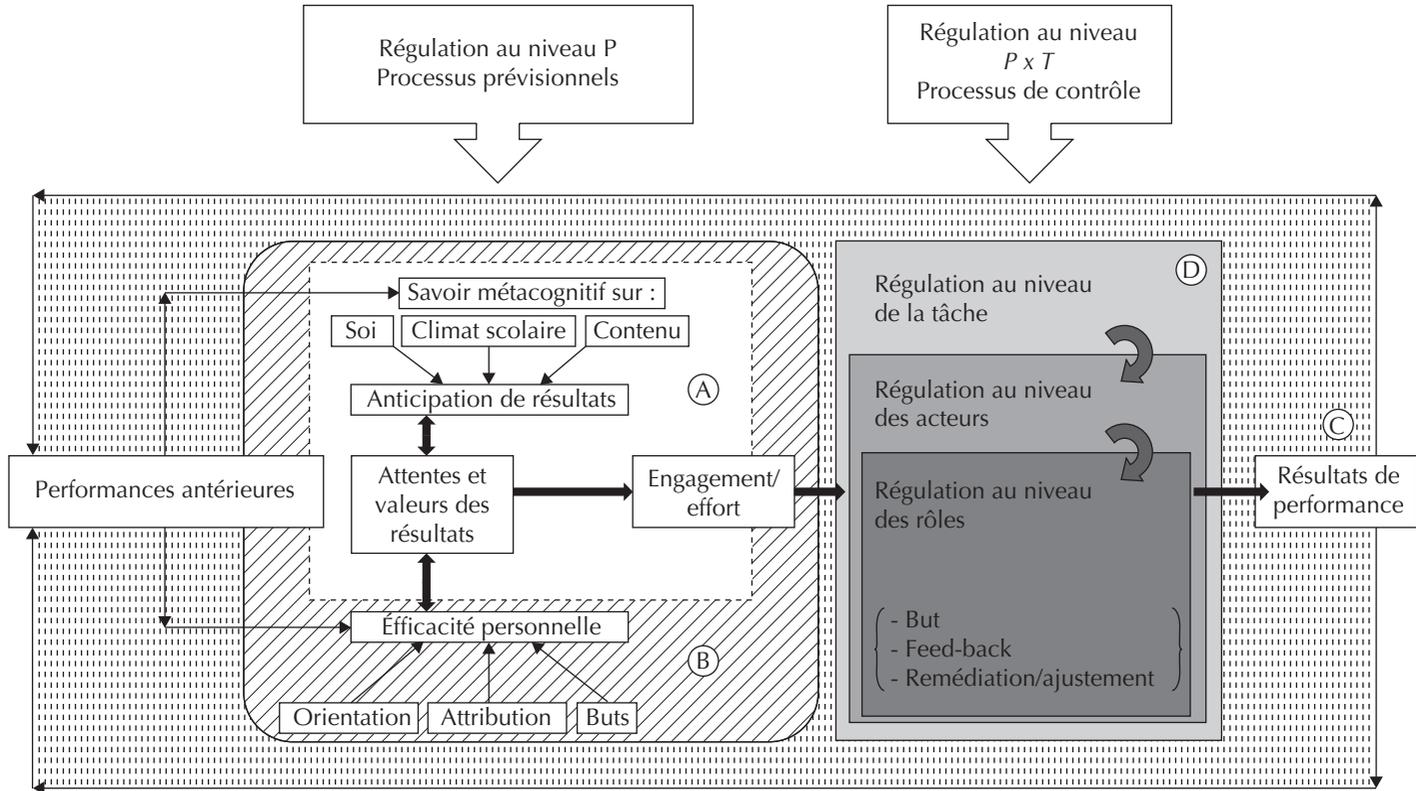
C'est afin de faire des élèves de meilleurs évaluateurs que l'EsA prend appui sur un modèle d'autorégulation des apprentissages. Ceci étant dit, il est difficile de concevoir comment il serait possible de faire des élèves de meilleurs évaluateurs sans développer de meilleurs modèles d'évaluation. Il faut comprendre par là non seulement les modèles d'évaluation formative qui informent l'enseignant et l'élève sur les résultats de ses apprentissages, mais aussi ceux qui contribuent à développer la capacité de l'élève à exercer un meilleur contrôle sur ses apprentissages, que ce soit à travers la régulation et l'autorégulation des processus cognitifs, métacognitifs, motivationnels ou émotionnels.

L'objectif de faire des élèves de meilleurs évaluateurs nécessite que nous comprenions mieux le fonctionnement des mécanismes d'autorégulation. Traditionnellement, les efforts de recherche se sont portés sur les mécanismes cognitifs et métacognitifs de la régulation. Plus récemment, des chercheurs se sont intéressés aux mécanismes motivationnels et émotionnels. Le rôle joué par les dispositions affectives et motivationnelles sur l'apprentissage des élèves ne se dément pas et, dans le cas des élèves en difficulté d'apprentissage, peut s'avérer particulièrement important. Le développement de modèles d'autorégulation intégrant l'ensemble des dispositions cognitives, métacognitives, motivationnelles et émotionnelles des élèves apparaît comme fort pertinent. Ceci étant dit, ces nouveaux modèles demeurent des chantiers de recherches fondamentales et appliquées largement ouverts et leurs implications pour l'évaluation demandent à être précisées.

MODÈLE EMPIRIQUE D'AUTORÉGULATION ET VALIDITÉ THÉORIQUE

La figure 1 propose un modèle faisant la synthèse des principaux facteurs intervenant dans l'autorégulation des apprentissages à partir des résultats de mes propres recherches. Ce modèle s'est construit par étapes successives dans le cadre de l'étude des facteurs psychologiques, principalement individuels, pouvant contribuer à expliquer la performance et aider à concevoir des modalités améliorées d'EsA.

Figure 1 : Modèle d'autorégulation illustrant l'évaluation-soutien d'apprentissage (EsA)



Les premières recherches ont porté sur l'impact des variables métacognitives sur la réussite à des examens (figure 1, section A). Laveault et Fournier (1990) démontrent qu'une partie importante de la réussite à un examen dépend de la capacité de l'élève à anticiper correctement les questions sur lesquelles portera l'examen, à s'autoévaluer et à ajuster sa préparation en conséquence. Des résultats similaires ont été obtenus en ce qui concerne la perception de la familiarité d'une tâche, les élèves s'avérant en mesure d'évaluer avec fidélité la familiarité de différents aspects d'une tâche (Jonnaert & Laveault, 1994), qu'il s'agisse des opérateurs, des objets ou des produits de la situation didactique. Les résultats de ces premières recherches illustrent bien le rôle actif que joue l'élève dans l'évaluation de ses apprentissages et l'importance du savoir métacognitif.

Ces premières recherches laissent cependant bien des questions en suspens. Pour quelles raisons un élève qui anticipe correctement les questions à un examen, ou encore qui perçoit nettement les caractéristiques d'une situation qui ne lui sont pas familières ne réussit-il pas ? Plusieurs explications d'ordre métacognitif ou motivationnel ont été envisagées. D'abord, il se peut que même en se représentant adéquatement une tâche, un élève ne s'y prépare pas ou ne mobilise pas toute l'attention nécessaire. L'engagement cognitif peut être affecté soit par le fait qu'il surestime sa préparation et ne porte pas à la tâche toute l'attention requise ou encore qu'il sous-estime sa capacité à accomplir la tâche avec succès et ne voit pas l'intérêt d'y investir les efforts nécessaires. Dans le premier cas, les facteurs à l'origine de l'engagement cognitif de l'élève sont d'origine métacognitive, dans le second cas, ils sont d'origine motivationnelle.

Les recherches subséquentes sur les représentations des élèves avant et après la réalisation d'une tâche complexe ont grandement contribué à préciser les interactions entre facteurs métacognitifs et motivationnels de la performance (Laveault, Leblanc & Leroux, 1999). Les résultats ont révélé que plus l'élève était confiant en sa réussite, plus il se fixait des objectifs élevés, mais aussi que plus la tâche lui paraissait claire, plus il était sûr de réussir. Les résultats ont également fait ressortir l'importance des régulations externes. C'est ainsi que les parents plus confiants que leurs enfants en leur réussite se sont avérés les plus soutenant dans l'aide aux devoirs (métacognition optimiste) : ils accordaient généralement plus d'autonomie à leurs enfants et manifestaient leur optimisme en leur fixant des cibles élevées d'apprentissage.

Dans la figure 1, la section A du modèle représentant le rôle du savoir métacognitif a été complétée ultérieurement par l'ajout de la section B qui comprend les facteurs motivationnels issus des théories sociocognitives : perception d'efficacité personnelle, nature des buts et orientation d'apprentissage, styles d'attribution. Le grand nombre de corrélations entre variables identifiées au départ comme métacognitives et motivationnelles a fait apparaître le besoin

de situer les recherches sur l'autoévaluation et de regrouper ces variables dans un ensemble plus vaste appelé *processus prévisionnels* qui comprend les régulations, tant cognitives que motivationnelles, intervenant sur la représentation que se fait l'élève des attentes de résultats et de leur valeur. Tout se déroule comme dans la fable de La Fontaine *Le renard et les raisins* : attentes et valeur des résultats sont étroitement associées. Le renard de la fable abandonne tout espoir et diminue la valeur de l'objectif – les raisins qu'il convoite – lorsqu'il ne parvient pas à les atteindre : « Ils [les raisins] sont trop verts et bons pour des goujats ». L'élève n'agit pas autrement lorsqu'il est confronté à des cibles d'apprentissage qui se situent trop souvent hors de sa zone proximale de développement.

Ces premiers résultats ont conduit à confirmer une distinction importante entre deux types de processus cognitifs d'autoévaluation (Paquay, Allal & Laveault, 1990) : une *évaluation par soi-même* « qui consiste à utiliser de manière relativement autonome une série de critères, internes ou externes pour évaluer » (p. 12) et une *évaluation de soi* qui possède un volet affectif et qui comprend le savoir métacognitif sur soi-même. Pour les deux types d'autoévaluation, l'étude du degré de confiance et du réalisme des évaluations des élèves s'est avérée un champ de recherche particulièrement fécond et utile.

En ce qui concerne l'évaluation par soi-même, le réalisme semble dépendre soit de la connaissance de la matière, soit de la connaissance des critères d'évaluation, soit des deux. Dans le premier cas, Boulé et Laveault (2011) démontrent que les étudiants qui performent le mieux expriment des degrés de confiance en leurs réponses qui sont davantage en accord avec leurs résultats. Il existe également un lien non linéaire (fonction quadratique en « U ») entre la difficulté d'une tâche et le réalisme de l'autoévaluation, celui-ci étant plus élevé sur les tâches nettement faciles ou difficiles, les valeurs de réalisme les plus basses se concentrant pour les tâches de difficulté intermédiaire. Laveault et Miles (2008) démontrent aussi que la précision avec laquelle les élèves évaluent le niveau de rendement de productions écrites est le meilleur prédicteur du degré de réussite de leur propre production. Dans les classes où les élèves se sont avérés les plus précis, ceux-ci se sont également avérés des juges plus fiables dans leur appréciation tel qu'attesté par des coefficients significatifs de concordance entre élèves. Ces résultats confirment que la précision de l'autoévaluation est influencée par le partage entre élèves d'une représentation commune des critères d'évaluation, d'où l'importance de développer les compétences de chacun à s'autoévaluer.

En ce qui concerne l'évaluation de soi-même, Laveault et Miles (2008) ont remarqué que le degré de précision des évaluations était affecté par le degré de sévérité de l'élève, certains élèves ayant une tendance systématique à surestimer le niveau de rendement, alors que d'autres ont

tendance à le sous-estimer. Sévérité et précision de l'autoévaluation sont par conséquent liées, les élèves trop ou pas assez sévères ayant tendance à errer davantage dans leur appréciation. Nous pouvons émettre l'hypothèse que ce manque de réalisme occasionné par des différences individuelles dans la sévérité des évaluations influence les processus prévisionnels de deux façons. D'une part, les élèves qui ont tendance à se surestimer ne feront pas les efforts ou ne s'engageront pas cognitivement dans la tâche à la hauteur des défis réels de celle-ci. D'autre part, les élèves qui ont tendance à se sous-estimer risquent de ne pas s'engager dans la tâche parce qu'ils considèrent comme très faibles leurs chances de réussite. Une telle hypothèse a été avancée par Laveault et Miles (2008) pour expliquer pourquoi les élèves identifiés comme doués et talentueux, pourtant plus précis que les élèves en difficulté d'apprentissage, n'étaient ni plus ni moins confiants en leur évaluation que les élèves faibles. Le degré de confiance serait modéré par la prise de conscience des difficultés de la tâche (*savoir métacognitif sur la tâche*) chez les élèves doués et par l'expérience personnelle de difficultés persistantes chez les élèves en difficulté d'apprentissage (*perception d'efficacité personnelle*).

Les processus prévisionnels ont une influence directe sur l'engagement cognitif et l'effort que l'élève mobilise. Ils regroupent les régulations intervenant avant l'exécution de la tâche et portant sur la personne de l'élève. Ils interviennent dans les processus d'autorégulation a priori portant sur l'autoévaluation que l'élève se fait de la tâche et de lui-même avant de l'entreprendre. Ce regroupement s'appuie sur l'hypothèse que l'autoévaluation est une capacité qui s'apprend graduellement et qui se caractérise par un meilleur contrôle exercé par l'élève sur ses processus. Cet apprentissage est représenté par une boucle de feed-back reliant l'autoévaluation a posteriori des résultats (figure 1, section C) à la représentation des performances antérieures. Cette boucle repose sur l'hypothèse que, tant le savoir métacognitif de l'élève que sa perception d'efficacité personnelle résultent de l'expérience cumulée des performances antérieures, chaque nouvelle autoévaluation alimentant ou pas des processus prévisionnels déjà établis.

Les sections A, B et C de la figure 1 regroupant les processus prévisionnels et d'autoévaluation a posteriori sont complétées par la section D qui décrit les processus d'autocontrôle assurant la régulation de l'interaction entre la *personne et la tâche* (PxT). Cette interaction PxT est représentée par un emboîtement de régulations intervenant à trois niveaux. Elle comprend les régulations de la tâche en cours d'exécution, celles-ci comprenant les régulations au niveau des acteurs, ces dernières incluant à leur tour des régulations au niveau des rôles (Laveault, 2007). Toutes ces régulations possèdent des objectifs, des boucles de feed-back et des ajustements (indiqués entre accolades) qui consistent à réduire l'écart qui sépare la situation de l'objectif visé (Ramaprasad, 1983) :

- La régulation de la tâche vise à réduire l'écart entre le niveau d'apprentissage initial de l'élève et le niveau visé. Le but est l'apprentissage.
- La régulation entre acteurs vise à rapprocher les conceptions qu'élèves et enseignants se font de la tâche, des critères d'évaluation et de leurs rôles respectifs. Le but est un meilleur fonctionnement des acteurs.
- La régulation des rôles vise à assigner aux acteurs des rôles qui leur conviennent. C'est ainsi que le rôle de l'évalué peut être joué non seulement par l'élève, mais aussi par un autre élève (évaluation mutuelle) ou par l'enseignant (*reverse feedback*). Il en va de même du rôle de tuteur. Le but est une meilleure représentation des rôles de chacun.

La section D du modèle fait intervenir des processus exécutifs de contrôle de l'activité influencés par les conditions extérieures. En milieu scolaire, élèves et enseignants partagent le contrôle de l'activité à différents degrés. Cet exercice du contrôle se traduit par le développement d'une capacité plus ou moins grande de l'élève à assumer le choix des buts d'apprentissage, du feed-back et de la remédiation (Laveault, 2007), trois composantes majeures de l'autorégulation. Par exemple, l'élève peut exercer un contrôle plus ou moins élevé des cibles d'apprentissage selon son degré de progression. De son côté, l'enseignant peut privilégier un contrôle plus grand sur le feed-back, lorsqu'un enseignement de précision s'avère plus approprié pour soutenir l'apprentissage de l'élève qu'un feed-back externe et détaillé, inutilement frustrant et contre-productif pour l'élève avec des difficultés sévères. Bref, le modèle implique que chaque niveau de régulation puisse faire intervenir des *microrégulations* au niveau des buts, du feed-back et de la remédiation qu'il s'agisse de la tâche, de l'acteur ou du rôle.

La section D du modèle permet de rendre compte aussi des situations d'évaluation mutuelle entre élèves, les rôles d'évaluateur et d'évalué pouvant être interchangeables en fonction des besoins. Dans l'étude de Laveault et Miles (2008), les élèves qui ont démontré les plus hautes performances se sont avérés les plus sévères, ce qui remet en question la pertinence d'associer des élèves forts avec des plus faibles dans des travaux d'équipe sans une préparation adéquate quant à la façon de réguler les rôles et les échanges entre acteurs. Dans une recherche inédite que nous venons de réaliser sur la communication du feed-back entre élèves, l'importance de mieux réguler les rôles et les échanges entre acteurs est apparue clairement. Dans une tâche où chaque élève, jouant le rôle de rédacteur de journal, devait informer un autre élève des motifs de refus de son article, la qualité du feed-back entre élèves est apparue fort variable. Visiblement, tous les élèves ne parviennent pas spontanément à fournir un feed-back constructif à un autre élève et cette capacité demande à être développée.

La validité théorique du modèle d'autorégulation de la figure 1 s'appuie en grande partie sur la théorie sociocognitive de l'apprentissage (Bandura,

1997) et réciproquement, plusieurs résultats des recherches ayant servi à élaborer le modèle confirment les principales hypothèses de cette théorie. La contribution majeure du modèle consiste à appliquer cette théorie à l'étude de l'EsA et au développement de pratiques qui développent les capacités de l'élève à s'autoréguler. Selon Butler et Winne (1995), pour qu'il y ait autorégulation, il faut une forme d'engagement de l'élève dans la tâche à accomplir, engagement au cours duquel l'élève exerce une série d'activités importantes :

- Déterminer le but d'apprentissage ;
- Planifier les activités à entreprendre ;
- Contrôler les activités en cours de réalisation ;
- Vérifier et ajuster l'activité en fonction de critères d'efficacité ou d'efficacité.

Le degré d'implication de l'élève dans les deux premières activités est pris en compte dans le modèle au niveau des processus prévisionnels (régulations a priori au niveau de la personne) et rend compte de la valeur perçue de la tâche et de l'engagement cognitif de l'élève. Quant à la troisième activité, le contrôle de l'élève au niveau de la réalisation de la tâche est décrit dans le modèle comme résultant d'une interaction $P \times T$ faisant intervenir un ensemble de régulations négociées au niveau de la tâche, des acteurs et de leurs rôles. La quatrième activité est illustrée dans le modèle par deux boucles de feed-back : la première consiste en une autoévaluation continue de l'exécution de la tâche et affecte directement les régulations au niveau de la tâche, des acteurs et de leurs rôles ; la seconde consiste en une autoévaluation a posteriori des résultats et contribue à plus long terme à ajuster les processus prévisionnels futurs, tant métacognitifs que motivationnels.

RÔLE DE LA RÉGULATION DES ÉMOTIONS DANS LES STRATÉGIES D'EsA

Les modèles d'autorégulation, incluant celui dérivé de nos propres recherches, sont pour la plupart des modèles cognitifs. L'absence des émotions en évaluation scolaire est un héritage de l'évaluation certificative où le caractère rigoureux et à visée objective tend à restreindre leur rôle au minimum. Ceci ne signifie pas que leur rôle n'est pas important. Tout au contraire, l'absence d'autorégulation des émotions telles que la colère, la tristesse ou l'anxiété peut faire en sorte qu'il sera difficile, voire impossible, pour un élève de porter attention à une tâche et au feed-back s'il éprouve de telles émotions.

L'étude de la régulation des émotions (RE) est récente, car plus de la moitié des publications répertoriées dans la base de données ERIC avant 2011 sur ce sujet ont été publiées au cours des cinq dernières années. Eisenberg *et al.*

(2001, cités par LaBillois & Lagacé-Séguin, 2009) définissent la RE comme une forme de contrôle qui s'exerce en fonction des buts à atteindre : « C'est le processus qui consiste à initier, soutenir, moduler ou changer la manifestation, l'intensité ou la durée de sentiments internes, de processus physiologiques liés à l'émotion et à ses aspects comportementaux concomitants (par exemple l'expression faciale) pour mieux atteindre les buts d'accomplissement »¹ (p. 306).

L'intérêt pour la RE répond à un besoin réel et plusieurs résultats de recherche justifient que l'on s'y intéresse. Gumora et Arsenio (2002) démontrent que les émotions négatives nuisent aux processus cognitifs en réduisant l'espace de la mémoire de travail et que les émotions positives contribuent à élargir le répertoire des idées et à générer davantage de stratégies. Fried (2011) est d'avis que les émotions peuvent véhiculer de l'information utile pour l'enseignant et améliorer les processus cognitifs. Graziano, Reavis, Keane et Calkins (2007) font également valoir qu'une RE inefficace réduit le contrôle de l'élève sur son comportement et affecte la relation maître-élève de manière négative.

Les stratégies de régulation des émotions (SRE) se développent durant la scolarisation et le modèle d'autorégulation porte à croire que l'EsA peut contribuer à faire en sorte que ce développement soit plus ou moins harmonieux. Pekrun et Stephens (2009) regroupent les SRE en quatre catégories :

- SRE portant directement sur l'émotion, telles que l'usage de techniques de relaxation ou de contrôle de soi ;
- SRE portant sur l'évaluation cognitive, telles que le recadrage des attributions causales consistant à revoir ses croyances sur le caractère plus moins stable ou définitif de facteurs à l'origine de la réussite ou des échecs ;
- SRE portant sur l'environnement académique, telles que des stratégies d'enseignement différencié ou d'apprentissage coopératif ;
- SRE portant sur la performance en agissant directement sur les stratégies d'étude et les prérequis qui peuvent exercer une influence sur l'amélioration du rendement.

Le modèle de la figure 1 permet d'envisager l'impact que certaines stratégies d'évaluation peuvent exercer sur l'apprentissage de régulations cognitives et aussi émotionnelles. Il permet d'expliquer en quoi des stratégies d'EsA telles que celles recommandées par Stiggins, Arter, Chappuis et Chappuis (2004) sont efficaces et déterminent ce que l'élève va faire en réponse aux trois questions suivantes qu'il se pose en rapport avec l'autorégulation : « Où est-ce que je vais ? » (buts) ; « Où suis-je rendu ? » (feed-back) ; « Comment m'y rendre ? » (remédiation). Voici quelques exemples de ces stratégies d'évaluation et de leur impact aux niveaux cognitif et émotionnel :

1. C'est nous qui traduisons.

- *Où est-ce que je vais ?* À ce niveau, les stratégies d'évaluation visent à fournir à l'élève une représentation claire de la cible d'apprentissage et à utiliser des copies types de modèles forts et faibles de productions. De telles stratégies sont de nature à accroître la perception de contrôle, à faciliter l'anticipation des résultats et le sentiment d'efficacité personnelle. En fournissant à l'avance des exemples de productions, l'élève peut se faire une idée plus juste de ce qui est attendu. Au plan cognitif, il peut se fixer des buts plus réalistes et recevoir un feed-back plus informatif. Au plan affectif, il peut avoir recours à des SRE lui permettant de choisir les aspects de la tâche qui se situent dans sa zone de confort.
- *Où suis-je rendu ?* Les stratégies d'évaluation à ce niveau portent sur l'usage régulier d'un feed-back descriptif et détaillé. Un feed-back régulier sur les stratégies et les processus figure, lorsque les mots justes sont employés, parmi les stratégies les plus efficaces pour soutenir l'élève dans la régulation de ses apprentissages. Le feed-back peut aider l'élève à utiliser des SRE qui consistent à réévaluer une tâche (en fonction du rôle qu'il y joue), à ajuster les cibles à des niveaux plus réalistes (en tenant compte des autres acteurs) ou à modifier son cadre attributionnel (en fonction des résultats de la tâche).
- *Comment m'y rendre ?* Stiggins *et al.* (2004) recommandent de concevoir des activités qui permettent à l'élève de se centrer sur un aspect de la qualité à la fois et de l'impliquer dans des pratiques réflexives, notamment en favorisant le suivi et le partage de ses résultats d'apprentissage. L'élève pourra ainsi réguler plus facilement la tâche lorsque des priorités sont identifiées et que le feed-back est ciblé sur les objectifs de départ. Au niveau cognitif, une telle approche évite la surcharge cognitive. Au plan affectif, elle rassure l'élève en faisant en sorte que sa réflexion porte sur son travail et se fasse à partir de normes internes. En axant sa réflexion sur sa propre progression et non sur une comparaison sociale et normative, cette stratégie favorise une orientation axée sur la maîtrise et une régulation des apprentissages plus constructive, dans ses dimensions cognitive et affective.

CONCLUSION

L'EsA relève d'intentions louables et généreuses : réguler les actions des élèves et des enseignants pour faire progresser les apprentissages des élèves et également pour faire des élèves des individus plus autonomes, possédant les compétences requises pour porter un jugement d'évaluation constructif sur leur performance et sur eux-mêmes. Des progrès importants ont été réalisés en ce domaine et ce, même si ces intentions n'ont pas toujours fait l'objet de recherches approfondies ou si elles n'ont pu être encadrées par des modèles théoriques forts.

L'autorégulation est l'un de ces modèles qui peut fournir un cadre théorique permettant de faire progresser l'EsA sur des bases scientifiques

valides. Les recherches empiriques sur l'autorégulation des apprentissages et, plus récemment, sur sa composante émotionnelle, contribuent à expliquer comment l'évaluation soutient l'apprentissage et permet de concevoir des stratégies prometteuses d'évaluation. Sur la base des observations qu'il nous a été donné de faire dans nos propres recherches et du modèle développé dans ce chapitre, quelques observations importantes se dégagent quant à un modèle d'EsA fondé sur l'autorégulation :

- L'efficacité de stratégies d'EsA dépend de la capacité de l'élève à mobiliser son attention sur les processus cognitifs et métacognitifs, et cette mobilisation est liée à la motivation et à la régulation des émotions. Pas de véritable métacognition sans un niveau adéquat de régulation motivationnelle et émotionnelle, celui-ci dépendant en partie des dispositions de départ de l'élève.
- Les interactions entre les facteurs cognitifs et affectifs de l'autorégulation figurant dans le modèle font qu'il est difficile de concevoir des stratégies d'EsA généralisables à tous les élèves. La réussite de certains élèves peut s'expliquer par leur capacité à réguler leur stratégie cognitive, alors que pour d'autres, elle s'explique par leur capacité à utiliser des SRE favorisant la persévérance et l'effort.
- L'EsA doit viser non seulement à améliorer la performance actuelle, mais aussi exploiter l'expérience métacognitive des apprentissages et des émotions afin de développer les capacités de l'élève à évaluer. Le but de l'EsA n'est pas seulement de faire en sorte que l'élève apprenne de ses erreurs, mais aussi qu'il ait l'occasion de célébrer ses réussites afin de développer ses propres compétences en matière de jugement d'évaluation.

Le modèle présenté dans cette contribution est un modèle individuel. L'environnement physique et social de l'élève, sa classe, son enseignant, ses pairs, le climat d'apprentissage n'y interviennent qu'indirectement à travers les régulations au niveau des acteurs et des rôles. C'est une limite importante. Par exemple, on peut se demander quel est le rôle de la famille sur les buts et l'orientation d'apprentissage et quelles SRE sont utilisées par les enfants pour résoudre des conflits cognitifs lorsque les orientations favorisées à l'école et à la maison sont différentes. Comment éviter des émotions débilitantes telles l'anxiété, la colère, le sentiment d'impuissance ou la perte d'estime de soi si l'élève ne parvient pas à concilier les différences entre valeurs de l'école et valeurs familiales, notamment lorsque les premières, axées sur la maîtrise, entrent en conflit avec les secondes, axées sur la performance et l'exigence de résultats qui le situeront parmi « les premiers » ?

Si l'intérêt de l'autorégulation, tant dans ses aspects cognitifs qu'affectifs, est de plus en plus acquis, les applications pratiques faisant suite aux recommandations issues de la recherche sont, dans bien des cas, des généralisations ou des inférences partiellement vérifiées. Les besoins sont de trois ordres :

- Un besoin théorique, qui consiste à introduire l'autorégulation dans les modèles actuels afin de les compléter ;
- Un besoin pratique, qui consiste à mobiliser les connaissances sur l'autorégulation dans des pratiques d'évaluation qui soutiennent l'apprentissage des élèves ;
- Un besoin de parcimonie, qui vise à simplifier les modèles actuels tout en conservant l'essentiel de leur utilité pratique.

Nous sommes loin de répondre aux besoins pratiques et aux besoins de parcimonie. Une question fondamentale se pose alors : l'introduction de plusieurs niveaux de régulation dans les modèles actuels est-elle en mesure de répondre aux exigences à la fois de la parcimonie scientifique et de l'optimisation des pratiques ? Nous estimons que c'est possible en rapprochant les explications théoriques actuelles de ce que le savoir d'expérience de la profession enseignante reconnaît comme faisant partie des meilleures pratiques en EsA. Il s'ensuit que, dans un tel contexte, *modèle* prend un autre sens. Il devient synonyme ici de *pratiques efficaces* au sens de *pratiques modèles* ou, si l'on préfère, de pratiques à reproduire dans l'EsA.

Le développement de modèles appliqués est nécessaire au développement de nouvelles stratégies d'évaluation et est lié à celui des modèles théoriques. Enfin, les deux types de modèles sont requis pour qu'apparaisse clairement comment faire progresser le jugement d'évaluation, tant chez l'élève que chez l'enseignant. Il ne suffit pas d'évaluer les compétences ; encore faut-il que les compétences des acteurs à évaluer évoluent de concert avec l'appropriation des connaissances et leur mobilisation en contexte.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. (2010). Assessment and the Regulation of Learning. In P. Peterson, E. Baker & B. McGaw (Éd.), *International Encyclopedia of Education* (vol. 3, pp. 348-352). Oxford : Elsevier.
- 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.
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York : W. H. Freeman and Co.
- Boulé, S. & Laveault, D. (2011). Utilisation du degré de certitude et du degré de réalisme dans un contexte d'évaluation diagnostique. In G. Raïche, K. Paquette-Côté & D. Magis (Éd.), *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation : Volume 2. L'évaluation* (pp. 31-48). Québec : Presses de l'Université du Québec.
- Butler, L. D. & Winne, P. H. (1995). Feed-back and self-regulated learning : A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.

- Cardinet, J. (1986). *Évaluation scolaire et mesure*. Bruxelles : De Boeck.
- Earl, L. (2003). *Assessment as learning : Using classroom assessment to maximize student learning*. Thousand Oaks : Corwin Press, Inc.
- Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M. et al. (2001). The regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*, 72, 1112-1134.
- Fried, L. (2011). Teaching teachers about emotion regulation in the classroom. *Australian Journal of Teacher Education*, 36(3), 117-127.
- Graziano, P. A., Reavis, R. D., Keane, S. P. & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology*, 45(1), 3-19.
- Gumora, G. & Arsenio, W. F. (2002). Emotionality, emotion regulation, and school performance in middle school children. *Journal of School Psychology*, 40(5), 395-413.
- Jonnaert, P. & Laveault, D. (1994). Évaluation de la familiarité de la tâche : quelle confiance accorder à la perception de l'élève. *Revue des sciences de l'éducation*, 20(2), 271-291.
- LaBillois, J. M. & Lagacé-Séguin, D. G. (2009). Does a good fit matter ? Exploring teaching styles, emotion regulation, and child anxiety in the classroom. *Early Child Development and Care*, 179(3), 303-315.
- Laveault, D. (2007). De la régulation au réglage : étude des dispositifs d'évaluation favorisant l'autorégulation des apprentissages. In L. Allal & L. Mottier Lopez (Éd.), *Régulation des apprentissages en situation scolaire et en formation* (pp. 207-234). Bruxelles : De Boeck.
- Laveault, D. (2010). Classroom Assessment in Policy Context (French Sources). In P. Peterson, E. Baker & B. McGaw (Éd.), *International Encyclopedia of Education* (vol. 3, pp. 432-437). Oxford : Elsevier.
- Laveault, D. & Fournier, C. (1990). L'évaluation par objectifs : une approche métacognitive. *Mesure et évaluation en éducation*, 13(1), 57-74.
- Laveault, D., Leblanc, R. & Leroux, J. (1999). Autoévaluation de l'apprentissage scolaire : interaction entre processus métacognitifs et déterminants de la motivation. In C. Depover & B. Noël (Éd.), *L'évaluation des compétences et des processus cognitifs* (pp. 81-98). Bruxelles : De Boeck.
- Laveault, D. & Miles, C. (2008). Utilité des échelles descriptives et différences individuelles dans l'autoévaluation de l'écrit. *Mesure et évaluation en éducation*, 31(1), 1-29.
- Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques*, 280, 47-64.
- Paquay, L., Allal, L. & Laveault, D. (1990). L'auto-évaluation en question(s) : propos pour un débat. *Mesure et évaluation en éducation*, 13(3), 5-26.
- Pekrun, R. & Stephens, E. J. (2009). Goals, emotions, and emotion regulation : Perspectives of the control-value theory. *Human Development*, 52(6), 357-365.

- Perrenoud, P. (1998). From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a Wider Conceptual Field. *Assessment in Education : Principles, Policy and Practice*, 5(1), 85-102.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13.
- Scallon, G. (2000). *L'évaluation formative*. St-Laurent, QC : Éditions du renouveau pédagogique.
- Stiggins, R. J. (2005). *Assessment For Learning Defined*. Assessment Training Institute. Consulté le 18 janvier 2012 dans <http://www.docstoc.com/docs/14728977/ASSESSMENT-FOR-LEARNING-DEFINED>
- Stiggins, R. J., Arter, J. A., Chappuis, J. & Chappuis, S. (2004). *Classroom Assessment for Student Learning. Doing it Right – Using it Well*. Portland, OR : Assessment Training Institute.

Les modes de pensée, organisateurs des modèles d'évaluation : pour la reconnaissance d'une épistémologie profane

**Michel Vial
Université de Provence**

INTRODUCTION

Le choix d'un modèle d'évaluation est, chez le praticien, une activité spécifique qui ne relève ni d'une prise de décision rationnelle, ni d'une épistémologie comparable à ce qui est attendu d'un chercheur. Des « modes de pensée » disponibles dans le social « agissent » l'évaluateur. C'est ce que la recherche clinique dont il est rendu compte ici veut expliciter. La mise au jour des liens entre ces « systèmes d'idées » et les modèles de l'évaluation peut devenir un des éléments de la formation à l'évaluation, pour permettre un choix raisonné du modèle d'évaluation en fonction du projet d'évaluation que l'on porte, et non plus seulement à partir de préférences « insues ».

LE CADRE THÉORIQUE

Dans la tradition pédagogique, le terme « modèle » est souvent utilisé pour désigner une formalisation généralisante, exemplaire : à partir d'une situation étudiée, on identifie ses invariants qui sont alors considérés comme des principes généraux permettant de reconnaître ou de préconiser d'autres situations de la même classe. On parle alors de « modélisation d'une situation ». Modéliser signifie dans ce cas-là exhiber une structure, c'est-à-dire une chaîne d'éléments liés par des relations stables (des invariants) enclenchés les uns à la suite des autres et produisant une fonction, réalisant une

transformation. Autrement dit, il s'agit d'une suite réglée d'enchaînements transformateurs. Ainsi, parle-t-on des modèles de la formation (Ferry, 1970, 1983). C'est une procédure d'objectivation de l'essentiel d'une situation. Le terme « dispositif » est alors plus exact que celui de « modèle » (Ardoino, 2007).

La modélisation dans la recherche en sciences humaines est un travail réalisé par le chercheur au cours de sa démarche, quelle que soit la méthode de recherche employée, pour *formaliser* par exemple ses résultats et les rendre communicables sous une forme ramassée : ainsi, parle-t-on du modèle d'apprentissage expérientiel de Kolb (1984).

Nous allons ici nous intéresser à un autre sens du terme « modèle », dans le cadre d'une psychosociologie de la relation éducative. Le modèle cette fois est interne au sujet, porté par lui, même à son insu. Quand ce sujet est un praticien évaluateur, on sait que son modèle, par exemple du produit attendu, interfère dans la correction avec sa volonté d'être objectif (Bonniol, 1981), puisque ce modèle idéal va s'infléchir pendant la correction en fonction de divers paramètres, dont celui de la faisabilité par ces élèves-là. D'où la préconisation de travailler à expliciter les critères avant de corriger, pour fixer le modèle du produit attendu et contrôler sa stabilité pendant la correction (évaluation formative [Genthon, 1983]). Puis est venue l'idée de travailler ces critères directement avec les élèves (évaluation formatrice [Deruaz, 1992]).

Par évaluation, on n'entendra pas seulement l'organisation de la validation des contenus d'apprentissage (le bilan, la certification, le sommatif, le pilotage, le contrôle du rendement), mais bien la mise en acte du projet de rendre intelligible ce qui est fait (Ardoino, 2000), donc la communication sur le rapport à la valeur ; le débat de valeurs (Schwartz, 2000) est initié par la question : « Dans ce qu'on fait, où est l'important ? » On part donc de l'hypothèse d'un lien entre la façon qu'a le sujet évaluateur de se tenir dans sa pratique et sa façon de penser le monde. Simultanément à l'adoption d'un modèle d'évaluation, il y a activation d'un « mode de pensée » (Vial, 2001 ; Vial & Mencacci, 2007). Le terme désigne des conceptions de la place de l'homme dans le monde, qui donnent des figurations différentes du sujet. Ce terme a été préféré à ceux, plus vagues de « systèmes d'idées », d'« approches » ou de « perspectives » que l'on rencontre dans la littérature, afin de marquer qu'il ne s'agit pas seulement d'un cadre de référence choisi et travaillé par le sujet, mais d'un ingrédient plus intime, avec lequel il se construit.

Dans ce contexte, ce que le chercheur travaille est l'organisation des modèles portés par les sujets enquêtés, leur exhibition formalisée, la mise au jour des références effacées mais mises en actes. Le modèle recueilli et rationalisé, didactisé (rendu enseignable) par le chercheur, est l'ensemble des principes d'action et des théories convoquées pour penser une pratique

qui la distingue d'une autre. C'est la matrice du dispositif qui sera, lui, une actualisation du modèle, permettant des variations dépendantes des conditions locales, concrètes, d'exercice. Le modèle d'évaluation est alors un ensemble organisé de principes d'action, dépendant chez le sujet évaluateur de « théories, souvent implicites, inconscientes, ou même contradictoires » (Sainsaulieu, 1987, p. 18).

MÉTHODOLOGIE DE LA RECHERCHE CONDUITE

Il a été conduit ici une recherche dans la méthodologie clinique, justement parce que le sujet est l'objet même de la recherche ; il fallait donc une « épistémologie qui fait place au corps et au sensible, à la temporalité du vivant » (Cifali & Giust-Desprairies, 2006, p. 10). La *singularité* du sujet social est précisément ce que la clinique étudie : la façon qu'a concrètement la personne d'organiser son monde et de s'y tenir. Comme l'affirme Dejours,

l'essentiel de ce qu'on cherche à évaluer se dérobe à l'observation directe. Pour y avoir accès, il faut recourir à des méthodes issues de la clinique, car les dimensions psychiques et intellectuelles du travail siègent dans l'expérience, à l'état de vécu, dans ce qu'on appelle aussi « expérience subjective du travail ». (2003, p. 13)

Il a été utilisé la méthode des cas qui consiste à relever ce qui, dans la particularité de chaque sujet, est exemplaire de la vision du monde des sujets aujourd'hui en éducation, pour exhiber un savoir

à la fois singulier et utile pour tous, c'est-à-dire pour tous ceux qui relèveraient du même cas qu'ils actualisent différemment. La psychanalyse nous a appris que relever d'un cas, n'est pas être dans la pure reproduction de ses symptômes, à l'identique, d'où le caractère transférable du savoir singulier par l'utilisation de taxonomies de symptômes jamais rencontrés pareillement agencés chez deux personnes. (Eymard, Vial & Thuilier, 2011, p. 83)

La liste des modes de pensée possibles joue ici le rôle d'une taxonomie de symptômes, une taxonomie permettant d'organiser les modèles d'évaluation.

Ce travail est une relecture du matériel d'une intervention faite en 2007, dans le cadre d'une expérimentation du portfolio en collège. Il s'agit d'une recherche-développement ayant pour but la mise en place d'un portfolio pour des élèves depuis la classe de 6^e jusqu'à la fin de la scolarité obligatoire (16 ans), dans un bassin d'enseignement secondaire français, suite à une commande du rectorat. Pour cela, l'option prise a été d'engager une recherche avec une équipe pluridisciplinaire d'une vingtaine de professeurs volontaires de cinq établissements du bassin. Le dispositif a articulé trois

phases : 2007-2008, élaboration du portfolio ; 2008-2009, expérimentations ; 2009-2010, essaimage pour d'autres élèves (Mencacci & Vial, 2009). L'évaluation est ici un processus continu greffé sur la pratique de l'accompagnement, à propos de l'utilisation du portfolio. Des entretiens en groupe avec les enseignants volontaires, pour expliciter comment ils mettent en place l'usage de ce portefeuille de compétences et l'accompagnement de l'élève ; des notes prises ; l'enregistrement audio des entretiens et des vidéos réalisées avec les élèves en constituent le matériau. Les discours des évaluateurs sont rapportés à cette grille qui sert à ordonner les items et déterminer les cas.

Tableau 1 : Les modes de pensée en évaluation

Modes de pensée, systèmes d'idées utilisés en évaluation	Façon de concevoir le destin de l'homme au monde. Figure du sujet. Lien avec un modèle d'évaluation.
Le déterminisme	<p>Les événements actuels ont avec les précédents une liaison fondée sur le principe évident qu'une chose ne peut commencer d'être sans une cause qui la produise (principe de la raison suffisante). Survvalorisation des rapports causaux. Les éléments sont des substances, des essences liées par les causalités. Le sujet est objet d'une sorte de fatum (il est agi par les lois). En cohérence avec le modèle de l'évaluation comme mesure.</p>
Le fonctionnalisme	<p>Se situer dans un schéma fins/moyens au nom de l'efficacité, à partir des théories de l'action rationnelle, dans la prise de décision rationnelle, la problématique. Principe de téléologie : téléguidé, l'acteur est motivé. Guidé par la fin, par le produit à réaliser ou la fonction à remplir, dans une trajectoire donnée pour atteindre la cible. Le sujet guidé est pensé comme un objet à fabriquer, dans la poïésis. En cohérence avec le modèle de l'évaluation par objectifs.</p>
Le structuralisme	<p>Mettre au jour des invariants, voire des universaux, où les éléments sont pris dans des enclenchements stables accomplissant des transformations, sur un sujet agi, effacé, dépendant de la structure. L'intérêt accru pour le fonctionnement provoque la recherche des mécanismes qui remplissent des fonctions (produisent des modifications). Le but est de voir la structure pour mieux fonctionner, ou de la réparer en corrigeant les dysfonctionnements. En cohérence avec le modèle de l'évaluation comme conseil pour l'aide à la décision.</p>

Modes de pensée, systèmes d'idées utilisés en évaluation	Façon de concevoir le destin de l'homme au monde. Figure du sujet. Lien avec un modèle d'évaluation.
La cybernétique	<p>Système fermé, coupé de son environnement, où prime la cohérence.</p> <p>Ensemble d'éléments en interaction, organisés en boucle, pour remplir la fonction du système.</p> <p>L'homme est un agent exécutant les ordres d'un chef lui-même soumis à un programme, un faiseur de produits calibrés, conformes à ce programme, qu'il faut surveiller pendant la réalisation des produits pour le régulariser de l'extérieur.</p> <p>En cohérence avec le modèle de l'évaluation formative.</p>
Le systémisme	<p>Système simple, ouvert sur un contexte avec lequel il forme un tout.</p> <p>Les interrelations sont privilégiées, par exemple : faire un avec la tâche en train de s'accomplir ; la tâche est l'environnement du sujet-système.</p> <p>Le sujet est un acteur qui interprète le texte des autres, qui s'adapte au milieu qu'on lui offre et trouve sa façon de résoudre le problème par des régulations successives.</p> <p>Le but est que le producteur prenne en charge la production commandée : il s'agit d'apprendre par l'autocontrôle. L'agent est responsable de ses réussites.</p> <p>En cohérence avec le modèle de l'évaluation formative.</p>
La systémique	<p>La complexité des systèmes : les systèmes de systèmes deviennent des « systèmes complexes », systèmes interconnectés, en grappes, multifonctionnels.</p> <p>Le sujet est en réseau, il est une connexion, le nœud d'un réseau de relations.</p> <p>Utilisation des paradoxes pour assurer la dynamique du système complexe.</p> <p>Survvalorisation de la « stratégie de la godille » : alterner les éléments antagonistes, utiliser les hiérarchies enchevêtrées du paradoxe ; faire une chose et son contraire.</p> <p>En cohérence avec le modèle de l'évaluation complexe.</p>
La complexité de l'humain	<p>Activation de la dialectique postmoderne et de l'herméneutique pour une praxis des situations fondée sur des contradictions irréductibles que la vie sociale demande pourtant d'articuler.</p> <p>Le sujet est un pluriel, un lecteur interprétant le monde comme un texte ; vivre c'est déchiffrer : la théorie est une fiction. Le sujet est produit social et producteur agissant sur le social ; divisé par l'inconscient, il n'est pas « maître chez lui », il interprète sans cesse : il évalue en continu, dans la dynamique des désirs.</p> <p>Le but est d'assumer les contradictions, de se situer dans le lien de contradiction pour rendre vivables les problèmes par la problématisation.</p> <p>En cohérence avec le modèle de l'évaluation située.</p>

Les matériaux apportés par les enseignants qui participaient à l'étude ont été examinés avec cette taxonomie qui est le résultat de recherches effectuées pendant plus de 10 ans (Bonniol & Vial, 1997 ; Vial, 2000, 2001 ; Vial & Mencacci, 2007 ; Vial & Thuilier, 2003). Pour identifier les éléments, il a fallu d'une part, faire une exégèse de textes sur l'évaluation et de compte rendus d'évaluations ; et d'autre part, faire des entretiens avec des concepteurs de dispositifs, avec des responsables d'évaluations et avec des évalués.

Moment considéré par la recherche

La recherche ici rapportée étudie les pratiques d'accompagnement. Il est important, à ce propos, de souligner que cette recherche considère un moment précis de l'histoire de la pratique de l'accompagnement chez les enseignants, soit avant que l'institution ne donne un cadre de référence en la matière, si confus soit-il, et alors que ce corps professionnel, contrairement au corps infirmier par exemple (voir Demailly, 2009), n'avait aucune expérience de l'accompagnement. Les enseignants ont toujours été soumis à l'injonction de guider l'élève, de contrôler les acquis, de vérifier la réalisation des programmes, de maîtriser les situations, la discipline, bref de piloter leur classe : l'accompagnement ne faisait pas partie de leur univers. Dans ce contexte, le travail institutionnel sur le livret de compétences et le socle commun de compétences a été un coup de tonnerre pour les enseignants à qui l'institution se mettait à dire qu'ils devraient « accompagner l'élève ». On est donc dans un moment spécifique, où un terme est introduit sans que les praticiens n'aient à disposition une conception et encore moins un modèle de référence. C'est un moment somme toute rare, où les modes de pensée deviennent particulièrement visibles. C'est pourquoi il a été choisi ici.

C'est le moment où l'on se retrouve dans le « magma d'une pratique » (Castoriadis, 1993) qui s'élabore en avançant, qui se spécifie peu à peu, ou bien sombre dans la confusion. C'est ce moment où la pratique est désignée sans effort de conceptualisation, par un terme ou une expression « qui est dans l'air » et qui fonctionne comme un « mot de la tribu ». Un mot émis par les institutions, lesquelles, par exemple, se mettent à parler de « gestion de projet », puis d'« accompagnement de projet », sans en donner une définition référée à un quelconque corpus théorique. Cette appellation est reprise et commentée, parfois sans être travaillée (la première définition qui passe, la plus ordinaire dans les usages quotidiens, est présentée comme la seule vraie) ; parfois aussi, elle est relative aux références convocables à un moment donné de l'histoire, et donc sujette à des effets de mode, de consensus implicites, de théories implicites. Ce peut même être des allants de soi : ainsi aujourd'hui encore, dans les institutions éducatives, le terme « accompagnement » recouvre des pratiques diverses, contradictoires, où le guidage n'est pas distingué de son contraire, l'accompagnement véritable

(ces pratiques ont peu de choses en commun, sinon le souci d'être bénéfiques à celui qui en est l'objet). « Accompagner » est alors utilisé comme un synonyme plus sophistiqué d'« aider » (Raucent, Verzat & Villeneuve, 2011). Ainsi peuvent apparaître simultanément plusieurs définitions qui coexistent et souvent se contredisent. Le même phénomène s'est produit avec les termes « contrat », « projet » puis « compétence » (Dolz & Ollagnier, 2002).

Processus de modélisation et choix des modèles d'évaluation

Plus la réflexion avance, plus les praticiens forment à une pratique donnée, plus les formateurs écrivent ce qu'ils transmettent, plus est donné à voir ce que doit être cette pratique, plus les chercheurs la formalisent, et plus les modèles deviennent visibles et la caractérisent, la distinguant petit à petit des pratiques voisines. Il n'y a donc pas d'abord une pratique de laquelle on extrairait les principes pour en construire le modèle, mais une mise en forme de la pratique, une modélisation progressive qui fait exister une pratique. C'est, plus tard, dans la théorisation que se donnent les praticiens réflexifs qui écrivent sur leurs pratiques et les chercheurs qui étudient les dispositifs mis en place, que s'installera le modèle valable pour un temps donné. Ainsi Hameline (1979), de Landsheere et de Landsheere (1982) en leur temps ont-ils contribué à propager le modèle de l'évaluation par objectifs, Allal (1979) le modèle de l'évaluation formative, Nunziati (1990) celui d'évaluation formatrice, etc.

Les modes de pensée influencent le choix des modèles d'évaluation. Ils ne sont pas propres à l'évaluation mais traversent tous les domaines des sciences humaines. C'est ce que Berthelot (1990) appelle des « schèmes d'intelligibilité » : « à l'aval des grands principes recteurs et en amont des constructions conceptuelles, un niveau spécifique de la construction de la connaissance » (p. 22). *Penser dans un modèle*, c'est utiliser un ensemble de principes, d'axiomes et de postulats, de théories choisies, qui ne sont visibles que parce qu'ils uniformisent les discours et les pratiques qui en découlent, chez ceux qui partagent le même mode de pensée. Par exemple, pour faire de l'évaluation « mesure d'impacts » (Graugnard & Heeren, 1999 ; Lenoir, 2006), il faut avoir survalorisé les relations de cause à effet, et participer du déterminisme, même si le déterminisme n'est pas nommé. En effet, l'évaluateur est la plupart du temps agi par le mode de pensée élu qui, généralement, est peu affiché et reste plus ou moins dissimulé. La théorie est une référence ouverte, le mode de pensée est très souvent une référence cachée, parce que le sujet y adhère. Sauf chez ceux, relativement rares, pour qui le mode de pensée est donné comme espace d'investigation théorique (comme les systèmes complexes chez Lerbet-Sérini [1999], par exemple). La théorie est signée : elle est l'œuvre d'un auteur. Le mode de

pensée et le modèle d'évaluation qui en découle sont, eux, anonymes, ils dépassent toujours ceux qui contribuent à les propager. Au contraire de la théorie, le modèle reste impersonnel.

LES RÉSULTATS

Les enseignants enquêtés ne couvrent pas l'ensemble de la grille présentée. Par exemple, aucun d'entre eux n'a eu recours au déterminisme et au modèle de la mesure. Trois cas sont donc présentés, sous forme de vignettes cliniques (technique couramment utilisée en psychologie), montrant trois types d'incidences du mode de pensée sur la conception du modèle d'évaluation, du portfolio et de l'accompagnement. En étudiant les cas, notamment dans la manière dont ils accompagnent spontanément leurs élèves dans la réalisation de leur portefeuille de compétences, nous avons pu les classer selon la grille ci-dessus.

Le cas d'Albert

Albert, la cinquantaine, enseigne les mathématiques. Son parcours est jalonné par les réformes successives de l'enseignement auxquelles il a, à chaque fois, pris part. Profondément investi de sa mission sociale, il veut faire réussir chacun de ses élèves au mieux de ses possibilités. Il les connaît et peut en parler des heures. Il tient un journal de classe, il a des tableaux de bord sophistiqués où sont consignés les tâches réalisées et les résultats de ses élèves. Il prend l'élève en charge, lui communique avant les apprentissages quels sont les produits attendus et les notions du programme utiles. Il essaie, dans la mesure du possible, de monter des séances de remédiation des erreurs, au moins sur les erreurs les plus fréquentes et les plus graves. Le portfolio de l'élève lui sert à communiquer ce qu'il a observé et à enregistrer les résultats. Il a interprété l'accompagnement, dans ces séances de rattrapage ou de soutien pour acquérir le programme, comme une communication pour que l'élève se situe par rapport à ce qu'on attend de lui et une remise à niveau. On est ici dans l'évaluation modélisée dans la cybernétique, avec le dispositif de l'évaluation formative, caractérisé par la rétroaction, la correction des erreurs pour une pédagogie de la réussite (Allal, 1979). L'accompagnement est dans ce cas une aide à la réussite des apprentissages et la construction d'un parcours sécurisé.

Le cas de Joëlle

Joëlle a la trentaine. Professeur de langues, elle met en place le repérage par compétences ; fière d'être pionnière, elle adhère pleinement à la réforme

en cours. Vive et brouillonne, elle virevolte et peut agacer ses collègues par le souci qu'elle a que ses élèves soient à l'aise dans sa classe, s'amusement presque. Sur les vidéos, on la voit au milieu de ses élèves comme une femme-orchestre, entrecoupant ce qu'elle a à enseigner, d'une apostrophe à l'un, d'un sourire de connivence à l'autre, d'une remontrance à un troisième. Quand elle analyse ce qu'elle fait, elle est débordée par la masse de données, de saynètes qui peuvent expliquer son travail et se perd dans l'anecdotique, mais elle clôt ses récits par des formules à l'emporte-pièce qui finissent par éclairer ce qu'elle sent. Elle pratique l'accompagnement du collectif classe, ce qui pour elle veut dire diriger, emmener le groupe où elle veut aller en usant de toutes les ruses possibles. Elle conduit sans cesse mais en étant tellement à l'écoute du groupe, tellement proche aussi, qu'elle dit faire de l'accompagnement. Elle écoute comment les élèves réagissent pour mieux les guider. Le portfolio est un cahier où l'élève s'essaye à des tâches qu'il a choisies dans une liste donnée par l'enseignante. On est ici dans l'évaluation modélisée dans la systémique, la complexité des systèmes : le dispositif de l'évaluation complexe, caractérisé par le tiers inclus, les hiérarchies enchevêtrées, la pensée conjonctive. L'accompagnement comme soutien utilise la « stratégie de la godille » : l'alternance entre du guidage et de l'accompagnement en fonction des besoins du parcours largement improvisé (Lerbet-Sérini, 1999), mais qui doit aller vers la réussite prévue.

Le cas de Roberte

Roberte est une jeune maman, professant le français, très calme. Elle parle peu mais acquiesce à ce qui se dit, puis pose une question qui oblige à faire la synthèse de tout ce qui a été formulé auparavant et qui semblait partir dans tous les sens. Elle saisit le moment favorable pour intervenir dans le questionnement en cours. En classe, elle lance des débats sans rester au premier plan. Les élèves participent beaucoup, parfois jusqu'au brouhaha ; elle rappelle alors le cadre du travail et les aiguille par ses questions ouvertes vers une production dont elle n'a pas du tout le détail en commençant. Elle se dit à l'écoute de leur progression et l'accompagnement sous forme d'entretiens individuels qu'elle organise avec le portfolio lui sert à approfondir la connaissance de tel élève qui vient demander de travailler tel point du programme, en dehors des cours. Le travail du portfolio ne consiste pas prioritairement à réaliser un produit, mais à créer une situation de dialogue, pour parler de ses difficultés et de ses joies, de l'utilité du savoir, des peurs d'échouer, de l'avenir aussi. L'important n'est pas de trouver des solutions, mais que l'élève arrive à mettre des mots sur ce qu'il ressent à propos du savoir, de son intérêt. Elle se situe plutôt du côté de l'affectif, du « psychologique », de l'existential, parce que ça lui convient de réhumaniser le rapport maître-élève et qu'elle pense que la majorité de ses collègues se situent de l'autre côté : l'acquisition du savoir ; qu'il y a

donc une demande chez les élèves et qu'elle ne perd pas de vue le savoir pour autant. On est ici dans l'évaluation modélisée dans la complexité de l'humain : le dispositif de l'évaluation située caractérisé par la problématisation et la création d'un lien de contradiction à explorer, où se situer par la dialectisation. L'accompagnement est vécu comme médiation (Mottier Lopez, 2006), pour permettre le cheminement de l'autre.

Le mode de pensée préférentiel de l'évaluateur le pousse à élire tel ou tel modèle d'évaluation. Ce mode de pensée n'est pas avoué comme tel : il est intériorisé. C'est la formation à l'évaluation qui peut donner aux praticiens les moyens de se repérer et donc de choisir le mode et le modèle qui en découle (Vial, 2012).

Le modèle crée l'objet

Personne n'a accès à la pratique, par exemple d'évaluation, sans être situé dans un modèle pour la penser. Par exemple, on n'a pas d'abord un objet appelé accompagnement puis une modélisation : le discours lit immédiatement la pratique à partir du modèle. Ce que nous ajoutons ici, c'est que la préférence, « insue » la plupart du temps, pour un mode de pensée incite à choisir tel ou tel autre modèle d'évaluation. Par exemple, parler d'évaluation située à propos de la pratique d'accompagnement, ce n'est pas chercher à situer l'évaluation dans la pratique d'accompagnement : elle y est partout, dans l'activité de l'accompagnateur, en tant que processus d'intelligibilité, sous la forme de la « réflexivité », comme dans tous les métiers de l'éducation, les « métiers de l'humain » (Cifali, 1994) ou des « ressources humaines » (Vial & Mencacci, 2007). L'évaluation située, comme les autres modèles, implique d'adhérer à certaines théories, à certains des concepts spécifiques qui la constituent comme modèle et donnent une version de la pratique de l'accompagnement qu'un autre modèle ne donnera pas. À partir de références inédites jusque-là dans l'évaluation en éducation, diverses adhésions sont repérables : à la dialectique postmoderne qui vise à assumer le lien de contradiction (Ardoino & De Perreti, 1998 ; Changeux & Ricœur, 1998), à la phénoménologie (Vermersch, 2004), à la psychanalyse (Giust-Desprairies, 2003), à l'herméneutique (Ricœur, 1986). Et l'évaluation modélisée dans la systémique utilisera, elle, pour penser l'accompagnement d'autres théories et arrivera à se donner d'autres principes pour agir.

Le modèle est l'application d'un ensemble de théorisations souvent implicites pour penser l'objet qui va dicter des principes d'action pour se tenir dans la pratique. En ce sens, tout modèle prescrit non pas d'abord ce qu'il faut faire, mais les éléments qu'il faut prendre en compte pour concevoir l'objet, la pratique. Différencier le modèle et l'objet, c'est installer une distance entre soi et l'objet qui permet le choix du modèle en fonction du projet que l'on porte et non plus seulement à partir d'une préférence personnelle : c'est le rôle de la formation à l'évaluation.

INTERPRÉTATION DES RÉSULTATS : POUR LA RECONNAISSANCE D'UNE ÉPISTÉMOLOGIE PROFANE

Les modes de pensée nous permettent d'identifier chez les praticiens une autre épistémologie que celle que l'on prête d'habitude aux chercheurs. On connaît l'existence et l'importance des savoirs « profanes » :

Du seul fait de vivre dans une société, toute personne a acquis du savoir sur cette société, aussi bien que sur elle-même. Ce savoir peut prendre différentes formes : il peut apparaître spontané, quasiment inconscient de lui-même, « insu » en quelque sorte, ou, au contraire, il peut avoir fait l'objet d'élaborations. Il peut être plus ou moins explicite, plus ou moins cohérent, plus ou moins pertinent. C'est dire que nous prenons ici le terme de *savoir* dans son acception la plus large. C'est en effet à partir de ce savoir que l'être humain se repère dans la société, dans la vie, et se confronte perpétuellement à des problèmes de valeurs et de sens. Il existe d'autre part, dans notre société, des « savoirs savants », ceux qui sont élaborés par des professionnels, payés pour produire des connaissances sur l'homme et la société. Ces savoirs puisent leur légitimation dans l'organisation de « cités scientifiques » où un contrôle collectif s'exerce sur la production de ces connaissances. En raison de la force d'autorité détenue par ces savoirs savants, nous nommerons les premiers, ceux du citoyen ordinaire, des « savoirs profanes ». (Feldman & Le Grand, 1996, p. 95)

Le praticien, avant d'avoir des modèles construits pour lire sa pratique, a des savoirs profanes, sous forme de schèmes, de modes, de registres, etc.

Si l'épistémologie est non pas seulement un domaine de la philosophie des sciences où se jouent les débats des savoirs savants (tradition francophone), mais le positionnement sur ce qu'est la connaissance, et donc ce à quoi le sujet donne statut de savoir (tradition anglophone), ces savoirs profanes peuvent aussi être regardés comme organisés dans des « architectures » (Pastré, 1999) ou des « constellations » (Barbier, 1996) qui leur confèrent alors le statut d'éléments appartenant à une épistémologie profane, puisque le sujet les vit comme étant de l'ordre du su, du vrai. Ils sont souvent appelés, ce qui peut être ambigu, « savoirs de l'action » (Barbier, 1996 ; Vergnaud, 1996), comme si leur origine était l'action (on risque alors de les confondre avec les habiletés, avec l'intelligence pratique [Vignaux, 2003]), alors qu'à l'évidence leur origine est sociale. Mais en revanche, ils servent bien l'action « crue bonne », sans passer par un quelconque processus de prise de décision rationnelle : ils s'imposent au sujet, comme des évidences. Les modes de pensée élus par les sujets relèvent d'un savoir profane chez eux ; quand ils sont construits et donnés à voir comme organisateurs des modèles d'évaluation, ils entrent dans les savoirs savants. Former à l'évaluation accompagne ce passage et la distanciation qu'il occasionne.

Comme Feldman et Le Grand (1996),

nous considérons *a priori* que savoirs profanes et savoirs savants ont chacun leurs propres légitimités, leurs propres limites, voire leurs propres erreurs et points aveugles. Cette position nous paraît particulièrement importante quand il s'agit de savoirs portant sur l'homme et la société : chacun a droit à ce savoir qui lui permet de se repérer dans la société où il vit. D'autre part, tout savoir savant portant sur l'homme et la société a des retombées sur chacun, directement ou à plus long terme. Il est donc essentiel de reconnaître que la dotation de sens n'appartient pas qu'au seul chercheur et que celui de toute personne a une validité et une autonomie qui ne saurait se confondre avec celle « en dernière instance » du premier. (p. 97)

D'autant plus que l'on peut aussi détecter chez les chercheurs cette même épistémologie profane qui se donne à voir sous la forme d'engouement pour un modèle plutôt qu'un autre, ou dans le rejet sans argument d'un modèle qui n'est pas le sien. Tous les chercheurs ne font pas le lien entre leurs objets de recherche, leur méthodologie et les modes de pensée : modéliser est un processus, ce n'est pas une simple procédure. Et comme tous les processus, il se vit ; il permet de penser. Il s'analyse après coup, quand il faut argumenter du résultat trouvé, quand il faut expliciter ce qu'on dit parce que l'autre ne comprend pas d'où l'on parle... Le processus se condense au gré de l'avancée du travail : il ne se restitue pas. La modélisation est un processus heuristique fait de méandres, de retours en arrière. La trajectoire n'y est pas de mise. Dans le contexte de la recherche, le concept pragmatique identifié (comme celui de « mode de pensée ») rassemble en lui tout le travail du processus. Les savoirs profanes intériorisés fonctionnent de façon idéologique tant qu'il n'y a pas eu de formation à l'identification de l'ensemble des modèles d'évaluation.

Comprendre d'où parle l'autre, c'est identifier la modélisation qu'il a faite et accepter que les modèles existants soient hétérogènes, incompatibles entre eux, jusqu'à un moment « t » d'une histoire existent plusieurs modélisations toutes utilisées. Chacune dit quelque chose de spécifique de l'objet, et le modèle est, en ce sens, un point de vue. La discussion scientifique ne peut donc pas invalider un modèle : tout dépend du projet que le sujet porte. Par exemple, l'évaluation comme mesure en éducation ne peut prétendre être la seule évaluation possible, mais si on veut contrôler le rendement, déterminer des effets et identifier des causes, si on pense l'évaluation depuis le mode de pensée du déterminisme, c'est le modèle le plus adéquat. Ce n'est pas le plus valide en soi, tout dépend de ce que l'on veut faire en évaluant. Le projet d'évaluer, avec ses variations, permet de rendre légitimes plusieurs modélisations et de débattre de leur utilité pour mieux comprendre les pratiques. Encore faut-il faire le deuil d'un modèle englobant tous les modèles et donnant une version unique de l'évaluation

en éducation¹. Se forcer à expliciter ses cadres de pensée et son ancrage théorique est donc utile. Pour autant, ce n'est pas restituer l'authentique d'un trajet, le processus est trop labile, c'est bien plutôt reconstruire une fiction. La modélisation est une aventure, on ne peut rationnellement en communiquer que des photos.

CONCLUSION

Le modèle en sciences humaines n'est pas une maquette sur laquelle on peut faire des expérimentations parce que l'objet, lui, n'est pas (ou pas aussi facilement) manipulable, comme dans les sciences de la nature. Nos objets sont des pratiques et non pas des substances. Il s'agit d'un modèle dit conceptuel, porté par des sujets qui, s'il est identifié à partir d'une ou de plusieurs théories élues, permet de se donner des principes d'action d'où vont dériver des dispositifs types. Le modèle sert de cadre préconstruit et il conceptualise la pratique qu'il organise, qu'il crée. La modélisation de l'évaluation dépend des modes de pensée que l'évaluateur porte et qui désignent des objets à prendre en compte. Le modèle est un patron, un gabarit, un mimétisme, une mise en forme de la pensée, un formatage. D'autant plus que les promoteurs des dispositifs eux-mêmes n'ont pas un savoir exhaustif sur les théories auxquelles ils se réfèrent (souvent sans le dire expressément) et n'en ont gardé qu'un trait ou deux qui leur ont « parlé ». Le mode de pensée auquel l'évaluateur se réfère, y compris sans le savoir, n'est la plupart du temps pas travaillé, il est vécu comme vrai. À partir d'un mode de pensée, le sujet invente un ensemble d'objets et des procédures qui caractérisent sa pratique et permettent de la modéliser. Le modèle est davantage qu'un angle de prise de vue sur l'objet étudié, c'est une mise en forme qui dépend du mode de pensée préféré, cette « conceptualisation tacite, en acte dans le langage ordinaire » (Clot, 1998, p. 148). Parler, par exemple, de la pratique de l'accompagnement pourrait inclure, ne serait-ce que par « hygiène scientifique », de dire à partir de quel mode de pensée on en parle et dans quel modèle de l'évaluation on se situe. Le mode de pensée influence le choix d'un modèle de l'évaluation, il est un élément organisateur de la pratique, un analyseur qui permet de donner sa place à une épistémologie profane.

1. Penser l'évaluation en éducation implique de mettre hors champ :

– L'évaluation des objets fabriqués dans le monde industriel, où l'on donne la primauté aux procédures quand la sécurité est confondue avec la qualité et qu'on multiplie les contrôles techniques.

– L'évaluation dans la finance internationale où sous l'effet de la mondialisation, on impose un modèle de la mesure des coûts, une surveillance des mises en actes des politiques publiques par des standards indiscutables appelés « normes ».

– L'évaluation médicale qui est un décalque de la méthode expérimentale, pour traiter les symptômes, où il s'agit de faire la preuve de l'efficacité du traitement.

Ces trois types d'évaluation obéissent à d'autres impératifs que l'évaluation en éducation.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. (1979). Stratégies d'évaluation formative : conception psychopédagogique et modalités d'application. In L. Allal, J. Cardinet & P. Perrenoud (Éd.), *L'évaluation formative dans un enseignement différencié* (pp. 129-156). Berne : Peter Lang.
- Ardoino, J. (2000). *Les avatars de l'éducation*. Paris : Presses Universitaires de France.
- Ardoino, J. (2007). L'articulation dispositif/situation. *Questions vives*, 8, 17-22.
- Ardoino, J. & De Perreti, A. (1998). *Penser l'hétérogène*. Paris : Desclée de Brouwer.
- Barbier, J.-M. (Éd.). (1996). *Savoirs théoriques et savoirs d'action*. Paris : Presses Universitaires de France.
- Berthelot, J.-M. (1990). *L'intelligence du social*. Paris : Presses Universitaires de France.
- Bonniol, J.-J. (1981). *Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires*. Thèse de doctorat en Lettres et Sciences humaines, Université de Bordeaux.
- Bonniol, J.-J. & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles : De Boeck.
- Castoriadis, C. (1993). Complexité, magmas, histoire. In M. Amiot, I. Billard & L. Brams (Éd.), *Système et paradoxe, autour de la pensée d'Yves Barel* (pp. 55-73). Paris : Seuil.
- Changeux, J.-P. & Ricœur, P. (1998). *Ce qui nous fait penser : la nature et la règle*. Paris : Odile Jacob.
- Cifali, M. (1994). *Le lien éducatif : contre-jour psychanalytique*. Paris : Presses Universitaires de France.
- Cifali, M. & Giust-Desprairies, F. (Éd.). (2006). *De la clinique, un engagement pour la formation et la recherche*. Bruxelles : De Boeck.
- Clot, Y. (1998). *Le travail sans l'homme ?* Paris : La Découverte.
- Dejours, C. (2003). *L'évaluation du travail à l'épreuve du réel*. Nancy : INRA éditions.
- De Landsheere, V. & de Landsheere, G. (1982). *Définir les objectifs de l'éducation*. Paris : Presses Universitaires de France.
- Demailly, L. (2009). Fortune et ambiguïtés de l'accompagnement. *Empan*, 74, 21-28.
- Deruaz, A.-M. (1992). L'évaluation formatrice, source de motivation. *Les cahiers pédagogiques*, 300, 39.
- Dolz, J. & Ollagnier, E. (2002). La notion de compétence : nécessité ou vogue éducative. In J. Dolz & E. Ollagnier (Éd.), *L'énigme de la compétence en éducation* (pp. 27-44). Bruxelles : De Boeck.
- Eymard, C., Vial, M. & Thuillier, O. (2011). *Le travail de fin d'études. S'initier à la recherche en soins infirmiers*. Paris : Lamarre.
- Feldman, J. & Le Grand, J.-L. (1996). Savoirs savants, savoirs profanes. In J. Feldman, J.-C. Filloux, B.-P. Lécuyer, M. Slez & M. Vicente (Éd.),

- Éthique, épistémologie et sciences de l'homme* (pp. 89-105). Paris : L'Harmattan.
- Ferry, G. (1970). *La pratique du travail en groupe*. Paris : Dunod.
- Ferry, G. (1983). *Le trajet de la formation*. Paris : Dunod.
- Genthon, M. (1983). *Évaluation formative et formation des élèves, effets de transfert des processus mis en œuvre*. Thèse de doctorat en Sciences de l'éducation, Université de Provence Aix-Marseille I.
- Giust-Desprairies, F. (2003). *L'imaginaire collectif*. Paris : ERES.
- Graugnard, G. & Heeren, N. (1999). *Guide méthodologique : l'évaluation de l'impact*. Lyon : F3E CIEDEL.
- Hameline, D. (1979). *Les objectifs pédagogiques en formation initiale et en formation continue*. Paris : Édition sociale française.
- Kolb, D. A. (1984). *Experiential learning*. Englewood Cliffs, NJ : Prentice-Hall.
- Lenoir, H. (2006). *Mesure d'impacts des dispositifs publics de formation*. Communication présentée à la 8^e biennale de l'éducation, INRP et Écoles Normales Supérieures, Lyon.
- Lerbet-Sérini, F. (1999). Guidance, compagnonnage et/ou accompagnement. *L'année de la recherche en éducation*, 145-177.
- Mencacci, N. & Vial, M. (2009). *L'autoévaluation des pratiques professionnelles à travers l'usage du portfolio au second degré du scolaire*. Communication présentée au 21^e colloque de l'ADMEE-Europe, Université catholique de Louvain.
- Mottier Lopez, L. (2006). Évaluation située des apprentissages : le rôle fondamental de la médiation sociale. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 82-89). Paris : L'Harmattan.
- Nunziati, G. (1990). Les objectifs d'une formation à/par l'évaluation formatrice. *Les cahiers pédagogiques*, 280, 48-64.
- Pastré, P. (1999). La conceptualisation dans l'action : bilan et perspectives. *Éducation permanente*, 139, 13-35.
- Raucent, B., Verzat, C. & Villeneuve, L. (2011). *Accompagner des étudiants*. Bruxelles : De Boeck.
- Ricœur, P. (1986). *Du texte à l'action, Essais d'herméneutique, II*. Paris : Seuil.
- Sainsaulieu, R. (1987). *Sociologie de l'organisation de l'entreprise*. Paris : Presses de la fondation nationale des sciences politiques et Dalloz.
- Schwartz, Y. (2000). *Le paradigme ergologique ou un métier de philosophe*. Toulouse : Octarès.
- Vergnaud, G. (1996). Au fond de l'action, la conceptualisation. In J.-M. Barbier (Éd.), *Savoirs théoriques et savoirs d'action* (pp. 275-292). Paris : Presses Universitaires de France.
- Vermersch, P. (2004). Aide à l'explicitation et retour réflexif. *Éducation permanente*, 160, 71-80.
- Vial, M. (2000). *Organiser la formation : le pari sur l'auto-évaluation*. Paris : L'Harmattan.

- Vial, M. (2001). *Se former pour évaluer, se donner une problématique et élaborer des concepts*. Bruxelles : De Boeck.
- Vial, M. (2012). *Se repérer dans les modèles de l'évaluation : histoire, méthodes, outils*. Bruxelles : De Boeck.
- Vial, M. & Mencacci, N. (2007). *L'accompagnement professionnel ? Méthode à l'usage des praticiens exerçant une fonction éducative*. Bruxelles : De Boeck.
- Vial, M. & Thuilier, O. (2003). *L'évaluation au quotidien des cadres de santé à l'hôpital*. Paris : Lamarre.
- Vignaux, G. (2003). La ruse, intelligence pratique. *Sciences humaines*, 137, 32-35.

Modélisation de la participation des acteurs dans le processus d'évaluation de la formation

Pedro Rodrigues* & Eusébio André Machado**

***Université de Lisbonne**

****Université Portucalense**

Dans cette contribution, nous présentons deux pistes de modélisation qui cherchent, à partir d'un ensemble de données empiriques, à construire une vision des logiques de participation des acteurs dans les processus d'évaluation en formation. Par conséquent, les travaux présentés ci-après serviront, dans un premier temps, de support pour la compréhension de l'évaluation des contextes de formation ainsi que du rôle que les sujets y assument. Les deux études citées ont été réalisées au Portugal. Dans un deuxième temps, nous chercherons à présenter un « discours de la méthode » qui fonctionne, sur un plan métathéorique, comme une réflexion sur le mode de construction d'une modélisation d'une réalité donnée.

UN CONCEPT MODÉLISATEUR : LE PROCESSUS DE PARTICIPATION INTÉGRÉ À L'ÉVALUATION

La conception et la mise en place des dispositifs et des mécanismes de participation à l'évaluation, tout comme leur description et leur analyse (ainsi que leur évaluation), exigeront de clarifier ce que l'on entend par « participation » et par « évaluation ». Ces deux notions imprègnent la culture et la vie des sociétés occidentales contemporaines, de telle sorte qu'elles semblent se banaliser, comme si elles renvoyaient à des réalités évidentes, qui ne seraient ni problématiques ni contestables bien que polysémiques (Figari, 2007).

De prime abord, la participation (c'est-à-dire l'acte de participer) est généralement comprise comme l'activité consistant à « prendre part à ». Les analyses de l'évaluation (De Ketele & Roegiers, 1999 ; Figari, 1996 ; Lesne, 1984 ; Stufflebeam & Shinkfield, 2007) permettent de situer la part qui revient à chacun dans l'ensemble de la chaîne des pratiques concernées par le processus d'évaluation. Or, en incluant le pilotage et la régulation de son implantation et, en amont, la détermination du référentiel d'évaluation, l'analyse est obligée de considérer la question de la gouvernance du dispositif d'évaluation et du pouvoir d'influencer ses décisions à tous les niveaux. Dans ce contexte, la participation acquiert inévitablement des contours politiques, éthiques et axiologiques et renvoie au concept de partage de pouvoir dans une acception démocratique (Lima, 1988, 1998).

L'évaluation, en tant qu'activité scientifique, formelle et intentionnelle, s'est développée dans le contexte plus vaste de l'émergence des sciences sociales et humaines au XIX^e siècle. Elle a hérité alors de présupposés épistémologiques qui exigeaient une séparation entre le sujet-évaluateur et l'objet évalué. Sous l'influence du paradigme positiviste dominant de la première moitié du XX^e siècle, cette dichotomie fondatrice a conforté le principe selon lequel l'activité d'évaluer serait un processus objectif qui permettrait de mesurer la performance, en recueillant des informations à l'usage exclusif du sujet-évaluateur (à savoir l'enseignant, l'école ou l'État).

Or, cette conception de l'évaluation, fondée sur la dichotomie fondatrice évaluateur/évalué, a été progressivement remise en question, surtout à partir des années 1960, d'une part pour des raisons de nature épistémologique, et, d'autre part, pour des motifs d'origine sociale et politique. En effet, sur le plan épistémologique, on assiste à l'affirmation de l'autonomie des sciences sociales et humaines qui, tout en remettant en cause le paradigme positiviste, réclament un statut unique pour l'évaluation en fonction de la complexité et de la singularité de son « objet ». Le principe de la séparation sujet/objet est abandonné, et des méthodes et techniques se sont développées dans le contexte de l'anthropologie sociologique et phénoménologique, en accentuant la participation des sujets évalués à leur propre interprétation de la réalité. Sur le plan politique et social, un processus d'*empowerment* (Fetterman, 2000) a été observé, selon lequel la production de connaissances ne serait pas seulement une activité cognitive, mais aussi une façon de comprendre et de transformer la réalité sociale.

Ainsi, la production de savoir et le changement social seraient des processus concomitants, dans lesquels les sujets peuvent surgir comme les protagonistes d'une logique émancipatrice sous certaines conditions. Dans ce cas, l'évaluation se fonde sur un modèle clairement actif et participatif qui exige plus de responsabilités, d'engagement, de prise de pouvoir, bref, plus d'action et de participation de la part des acteurs : « l'évaluation et l'action

font partie intégrante du même processus au lieu d'être deux processus séparés dans le temps » (Guba & Lincoln, 1989, p. 259).

Dans le domaine de la formation, cette révolution copernicienne s'est traduite par une reconfiguration radicale des dispositifs d'évaluation qui ont été progressivement dominés par une logique plus globale, cyclique et itérative avec l'introduction de « l'évaluation des besoins de formation », « l'évaluation formatrice » et, en particulier, les processus « d'autoévaluation ». La principale caractéristique de cette révolution est la tentative de focaliser l'évaluation sur les sujets évalués, autrefois considérés comme de simples « objets », mais désormais envisagés comme des agents privilégiés dans le processus d'évaluation et, par conséquent, comme protagonistes dans la construction des dispositifs de formation eux-mêmes.

PREMIER TRAJET DE MODÉLISATION : DISPOSITIF INFORMEL D'ÉVALUATION

La modélisation présentée ci-après dérive d'un modèle général d'analyse qui vient de la description/conceptualisation d'un ensemble diversifié de pratiques intégrant les domaines de l'éducation et de la formation, parmi lesquelles figurent les pratiques d'évaluation et, en particulier, l'évaluation de la formation (Machado, 2007 ; Rodrigues, 2001, 2004, 2006). Ce modèle est confronté à l'évaluation réalisée par les formés lors d'un entretien individuel dans le cadre d'une formation continue destinée à des enseignants.

La modélisation de départ : les modèles d'analyse qui encadraient la recherche

Les analyses éthiques et politiques dans le domaine de l'évaluation (Greene, 2006 ; House & Howe, 2003 ; Karlsson & Conner, 2006) examinent la question du pouvoir et de son exercice. Ces analyses s'étendent à tous les secteurs de l'action sociale, y compris à l'éducation et à la formation, ainsi qu'à la recherche scientifique et à la science elle-même (House, 1990), en configurant des « paradigmes d'évaluation ». Elles permettent de situer des positions, des statuts, des rôles et des pouvoirs dans ce domaine, ainsi que de typifier et de graduer les pratiques et les modes de participation des acteurs (voir tableau 1).

Sous l'égide de l'objectivisme positiviste, l'autorité s'impose, légitimée par un savoir éprouvé et incontestable : il n'y a pas d'arguments contre les faits. Associée à des processus techniques et scientifiques, la recherche de la vérité surpasse les opinions et les perspectives particulières. Cette vision de l'évaluation implique, par conséquent, une division claire des statuts, des rôles et des pouvoirs de chacun, ainsi qu'une relation fortement asymétrique entre les différents intéressés et participants à l'évaluation.

Les perspectives subjectivistes qui mettent l'accent sur l'individu et qui reconnaissent sa capacité de compréhension et de liberté d'action, comme la phénoménologie, ont tendance à lui attribuer un rôle important dans la détermination de la formation et de l'évaluation. Pour leur part, les perspectives qui, telle la théorie critique, insistent davantage sur l'organisation et les contextes sociaux, et sont caractérisées par des jeux de pouvoir et des conflits d'intérêts, soulèvent des questions qui peuvent ouvrir la porte aux perspectives dialogiques et dialectiques d'analyse et d'intervention, y compris dans le domaine de l'évaluation.

Tableau 1 : Les paradigmes de l'évaluation comme paradigmes de participation

Dimensions	Paradigme d'évaluation		
	<i>Objectiviste</i>	<i>Subjectiviste</i>	<i>Dialectique</i>
<i>Épistémologie</i>	Positivisme	Phénoménologie	Théorie critique
<i>Axiologie</i>	Objectiviste	Relativiste	Dialectique
<i>Politique</i>	Autocratie	Suffragiste	Critique sociale
<i>Éthique</i>	Autoritaire	Non imposé	Consensuelle
<i>Source du pouvoir</i>	Hiérarchique	Individuelle	Communautaire
<i>Concept légitimateur</i>	Réalité du pouvoir	Liberté individuelle	Participation
<i>Type de participation</i>	Décision des autorités	Choix, vote, expression d'opinion	Négociation
<i>Statut de l'enseignant/formé/évalué</i>	Objet d'évaluation et source d'information	Client, évaluateur	Participant, coévaluateur
<i>Mode de travail de l'évaluation</i>	Non participatif	Participatif de type individuel et indépendant	Métaparticipatif de type dialogique et collectif

Entre le faible pouvoir des acteurs (sous le joug de l'objectivisme) et leur apparent pouvoir non questionnable et incoercible (dans la cosmologie individualiste subjectiviste), il peut y avoir différentes configurations participatives (aussi bien individuelles que communautaires). La plus large est celle qui s'étend à la planification et à la régulation conjointes (par les différents acteurs concernés) du processus d'évaluation lui-même, qui, étant participatif sur ce plan, inclura, à la limite, un niveau de métaparticipation.

Modélisation de l'évaluation participative de la formation par les enseignants formés

L'analyse que nous présentons ici porte sur l'évaluation réalisée lors d'un entretien individuel par 18 enseignants formés, dont 10 qui ont effectué une post-graduation universitaire d'un an et demi et 8 qui ont participé, par ailleurs, à un projet de recherche-action-formation pendant trois ans.

Du point de vue de la relation à l'objet d'appréciation, cette évaluation ne configurera pas une fonction formative et de régulation. En outre, le chercheur/intervieweur n'est pas mandaté pour réaliser une évaluation et, d'ailleurs, il ne l'entreprend pas. De ce fait, les enseignants n'effectuent pas une évaluation de la formation dans le cadre d'un dispositif formel. Cependant, nous émettons l'hypothèse que cette situation d'entretien est comparable à celle que l'on rencontre dans des contextes où les apprenants sont invités à se prononcer (en remplissant des questionnaires d'évaluation de la formation et/ou de l'enseignement, par exemple) tout en ignorant le sort qui sera donné à leurs réflexions et à leurs opinions.

L'évaluation est réalisée par chacun individuellement, en présence d'un interlocuteur qui n'est là que pour enregistrer le monologue et la réflexion de la personne interrogée. Le risque est qu'il n'y ait pas de dialogue ni de discussion, comme ce serait le cas s'il s'agissait d'une enquête par questionnaire.

En ce qui concerne la concrétisation de ces dispositifs, qui exaltent l'individualisme, il semble que la participation soit conçue comme un choix personnel (en situation « d'égalité » : « une personne, une évaluation », un client, un électeur). Sans doute qu'un des facteurs qui conditionnent la mise en place des dispositifs de participation incluant l'interaction directe entre les participants est l'ampleur du système qui est l'objet de l'évaluation (Stake & Schwandt, 2006). S'il est trop large et très diversifié, il sera difficile de construire des dispositifs pour offrir les conditions d'un face à face à tous les acteurs.

Il est évident que nombre de questionnaires, qui se limitent à demander une appréciation quantitative sur chaque élément/aspect de la formation considérée, ne permettent pas d'identifier les valeurs et/ou critères mobilisés dans l'évaluation ni d'interpréter les jugements portés (Pique, 1990). L'entretien par contre a l'avantage de permettre d'explicitier les appréciations et leurs fondements, tout comme le travail (cognitif) d'élaboration d'un jugement.

L'étude de l'évaluation réalisée par notre groupe d'enseignants révèle cependant que cette dernière a été en mesure de soutenir une activité dialogique et dialectique (interne) qui confronte les intérêts, les priorités et les valeurs de chacun, aussi bien avec ceux des autres qu'avec ceux des organisations en présence (y compris de la formation). Mais les participants

confrontent également des modèles (idéaux) de formation alternatifs et même antagonistes (voir tableau 2) et leur concrétisation spécifique (expérimentée), en prenant en compte les conditions et contraintes vécues ainsi que les réalisations qu'ils croient possibles dans ces circonstances-là.

Tableau 2 : Dimensions privilégiées de la formation évaluée par les participants

N°	Formation traditionnelle formelle (Post-graduation universitaire)	Formation professionnelle non formelle (Recherche-action-formation à l'école)
1	Déterminée par des besoins généraux	Enracinée dans des besoins personnels
2	Ancrée dans le savoir scientifique	Fondée sur les problèmes de travail
3	Basée sur la théorie	Centrée sur la pratique et l'expérience
4	Fournissant une information organisée	Exigeant la construction d'un savoir
5	Structurée d'avance	En évolution
6	Destinée à l'individu	En fonction du groupe (local) de travail
7	Nécessitant la lecture de textes	Exigeant (auto-)observation et auto-analyse
8	Développant l'argumentation	Opérant une transformation interne
9	Avec un calendrier fixé	Au rythme du groupe
10	Certifiante	Promouvant une meilleure performance

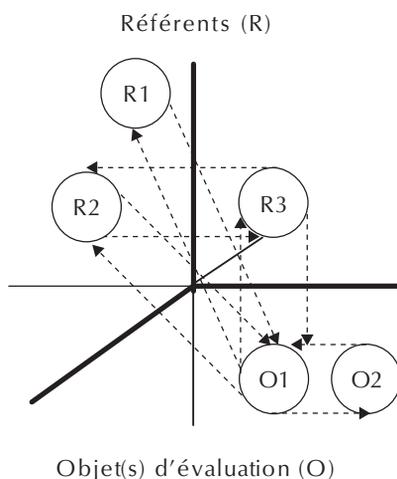
Ce tableau, ou cette analyse, montre que les enseignants décrivent les actions de formation auxquelles ils ont participé selon un ensemble de dimensions qui leur permet de les caractériser au moyen de descripteurs désignant des qualités de l'objet d'évaluation ; celles-ci acquièrent le statut de valeurs lorsque les participants+ leur attribuent de l'importance. Ils le font en typifiant des modèles de formation dont ces actions relèvent. Cependant, ces caractéristiques de la formation, bien qu'apparemment en opposition, s'avèrent également valorisées et souhaitées par les formés. Chacune des actions de formation avait ses points forts et ses points faibles, les points forts de l'une étant les points faibles de l'autre, et vice versa.

Après la réalisation de leur « mémoire » de post-graduation, certains enseignants ont indiqué que ce travail, tout comme la recherche-action,

leur avait fait développer une attitude active de recherche d'information et de solutions, contrairement aux actions de formation auxquelles ils étaient habitués. Autrement dit, leur façon d'envisager la formation (les référents mobilisés et l'évaluation) avait changé.

La participation individuelle des enseignants à l'évaluation de la formation semble relever de processus homologues aux processus de participation collective et de négociation entre les différentes parties intéressées ; ces processus conçoivent la participation comme un travail de consensus, après confrontation entre valeurs et référents, entre référés, puis entre les uns et les autres (voir figure 1).

Figure 1 : L'évaluation en tant qu'interaction entre le référentiel et l'(es) objet(s) d'évaluation



Dans cette figure, R1, R2 et R3 configurent des « types idéaux » de formation qui se confrontent sur le plan théorique/idéal ; O1 et O2 représentent des actions de formation concrètes qui peuvent être comparées non seulement les unes aux autres en termes de réalisation et d'expérience personnelle, mais aussi avec les idéaux auxquels elles se réfèrent. Dans un espace défini par les dimensions (ou critères) d'évaluation et peuplé de « modèles théoriques » d'une part, et d'« objets concrets » d'autre part, l'évaluation implique davantage un va-et-vient entre les différents plans successifs qu'un travail linéaire, unidirectionnel, déductif et descendant, sous un référentiel fixe.

Autrement dit, l'évaluation informelle individuelle, bien qu'autonome, n'est pas (complètement) indépendante et dérive d'un dialogue interne qui reproduit et actualise des débats sociaux entre divers points de vue,

qui, comme tels, sont considérés et examinés de façon critique. Ce n'est pas non plus un processus linéaire vérificateur ; il s'agit d'une démarche interactive de (re)construction réciproque du référentiel, des référents, de la représentation des référés, de l'expérience personnelle et de l'évaluation même, en tant que construction de sens.

DEUXIÈME TRAJET DE MODÉLISATION : DISPOSITIF FORMEL D'ÉVALUATION PARTICIPATIVE

La modélisation présentée ci-après prétend être, plus qu'une lecture normative, une description « des faits » et des pratiques d'évaluation tenant compte de la question transversale de la participation des sujets. En même temps, elle constitue un ensemble d'hypothèses de travail à partir desquelles on a cherché une possible consolidation empirique. Or, comme tout processus de modélisation, il s'agit d'une construction « idéale » qui, tout en offrant la possibilité de produire une abstraction qui rationaliserait ce qui existe, n'échappe pas au risque d'éliminer la complexité et la richesse empirique que l'on prétend étudier ici.

La modélisation comme point de départ : un référentiel pour la recherche

Cette deuxième modélisation de la participation à l'évaluation s'inspire surtout, d'un point de vue formel, du « modèle d'analyse des pratiques de formation » de Lesne (1984), bien qu'elle réunisse divers apports théoriques et métathéoriques (Bonnioi & Vial, 1997 ; Rodrigues, 1994, 2002, 2006 ; Scriven, 2000). Nous distinguerons ainsi trois procédures de travail évaluatif (PTE) qui, bien que séparées ici, ne doivent pas être considérées hiérarchiquement ni selon une logique de valorisation. Il faut, en revanche, prendre en compte les relations de contiguïté et de perméabilité qui s'établissent entre elles.

Tableau 3 : Une modélisation des pratiques d'évaluation dans la formation

Dimensions	Procédures de travail évaluatif		
	<i>PTE 1 Non participative type externe hétéronome</i>	<i>PTE 2 Participative type interne indépendante</i>	<i>PTE 3 Métaparticipative type négocié dialogique</i>
<i>Finalités</i>	Contrôler, vérifier, mesurer	S'épanouir	Réguler, coopérer, appuyer, autonomiser, transformer

Dimensions	Procédures de travail évaluatif		
	<i>PTE 1 Non participative type externe hétéronome</i>	<i>PTE 2 Participative type interne indépendante</i>	<i>PTE 3 Métaparticipative type négocié dialogique</i>
<i>Contenu</i>	Output Produit	Processus Contexte Induit/construit/ produit	Processus/produit Contexte Construit/produit
<i>Prise de décision</i>	Autocratique Bureaucratique Externe Non négociée	Adhocratique (Mintzberg, 2010) Interne	Démocratique Négociée Contractuelle
<i>Rôle de l'évaluateur</i>	Externe Formel	Interne Informel	Externe Coopération et/ou engagement
<i>Modalités d'évaluation</i>	Hétéro-évaluation Quantitative	Autoévaluation Qualitative	Interévaluation Quantitative- qualitative
<i>Paradigme d'évaluation</i>	Objectiviste	Subjectiviste	Critique

Voyons donc de quelle manière on peut caractériser chacune des PTE :

PTE 1, non participative de type externe hétéronome : ses finalités principales sont de contrôler, vérifier et mesurer, et relèvent donc d'un paradigme psychométrique et docimologique de l'évaluation ; elle part également d'un principe épistémologique de nature positiviste, selon lequel le sujet et l'objet doivent appartenir aux sphères mutuellement transcendantales (idéologie séparatiste). C'est pourquoi l'évaluateur adopte une position externe et formelle, garantissant ainsi l'objectivité des jugements formulés et l'autorité dans la prise de décision.

En raison de sa légitimité scientifique, l'évaluateur n'a pas besoin de négocier ou de dialoguer, puisque ses décisions peuvent assumer des caractéristiques à la fois autocratiques et bureaucratiques. En se limitant aux produits du processus d'évaluation, la PTE 1 se réalise à partir des modalités d'hétéro-évaluation de nature quantitative et sommative.

PTE 2, participative de type interne indépendante : il s'agit ici d'un ensemble de procédures qui, en dernière instance, cherchent à radicaliser le rôle des sujets évalués. De ce point de vue, outre la participation des sujets, il s'agit de faire en sorte que ceux-ci soient les auteurs mêmes du processus d'évaluation dans un but d'indépendance et d'épanouissement personnel. L'évaluation se règle par des critères, notamment en raison de

la singularité et de l'incommensurabilité des contextes. À la limite, ces procédures tendent à la prise de décision uniquement par autoréférence du fait que l'évaluateur lui-même ne se différencie pas du sujet évalué et qu'il assume un rôle interne et informel.

Les modalités d'évaluation les plus privilégiées appartiennent essentiellement à l'autoévaluation de type qualitatif et autocompréhensif. Les risques potentiels de ces procédures, qui relèvent d'un paradigme subjectiviste, résident dans l'absence de dialogue et de conflits (sauf ceux du sujet avec lui-même), ce qui pourrait ramener à des positions conservatrices et réductrices comme celles que l'on trouve dans un paradigme objectiviste.

PTE 3, métaparticipative de type négocié dialogique : la procédure vise à établir l'évaluation comme un processus de régulation, de soutien et de coopération, dans la mesure où elle valorise essentiellement les aspects liés aux politiques de prises de décisions à partir d'un référent démocratique et pluraliste. De ce point de vue, on insiste sur le dialogue social en tant que pratique sociale à la base des décisions et des contrats (contractualisation). D'un point de vue externe, l'évaluateur ne manque pas de s'engager et de collaborer avec « l'objet » sur lequel il prétend fonder son jugement, remettant ainsi en question le présupposé positiviste de la séparation entre le sujet et l'objet. Il valorise non seulement le produit mais aussi le processus.

Ainsi, l'évaluation provient de la participation des sujets, de la négociation des référents, de la définition de critères, de la production des jugements et des prises de décisions. Il s'agit donc d'une logique d'« interévaluation » avec une tendance strictement formative qui peut être considérée sous l'angle d'un paradigme critique, dans lequel l'évaluation surgit comme une coconstruction et une relation dialectique (Rodrigues, 2002).

Modélisation de la participation des enseignants formés à l'évaluation de la formation

L'analyse des données recueillies auprès d'un groupe d'enseignants par questionnaire (n = 221) et par entretiens (n = 12) révèle l'usage de pratiques de la « non-participation » ou de « participation atténuée », dans lesquelles les formés sont plutôt vus comme destinataires et non comme participants actifs dans la construction des dispositifs de formation et dans le processus d'évaluation. En effet, la participation des enseignants, dans la formation en général et dans l'évaluation en particulier, est faible, inexistante même dans des moments critiques et ne revêt donc pas beaucoup d'importance, que ce soit à un niveau quantitatif ou qualitatif. Nous sommes là face à ce que l'on désigne généralement par « participation mitigée ».

Ainsi, malgré l'existence d'une politique de formation centrée sur l'autonomie, la décentralisation et, plus particulièrement, sur la participation des enseignants et de l'école, les résultats confirment une faible participation ou même l'absence totale de participation des enseignants à l'évaluation et à la construction des dispositifs de formation.

Les enseignants sont considérés par les centres de formation comme des clients d'un marché qui prétend apporter un service et, de leur côté, ils assument le rôle de clients qui demandent un service. Par ailleurs, quand on mène une réflexion sur la participation des enseignants à l'évaluation de la formation, on doit tenir compte du fait que l'on ne peut pas participer quand on n'est pas (ou alors quand on ne se sent pas directement) impliqué ou considéré comme un partenaire. Dans le meilleur des cas, les enseignants sollicitent ou suggèrent des thèmes, des contenus et des actions, sans un travail de médiation réflexive ni une collaboration associative entre les « besoins » ressentis et les solutions formatives proposées.

En confrontant la modélisation que nous avons présentée (tableau 3) avec la réalité étudiée, il est possible de dégager les conclusions suivantes :

- Le début de la formation est considéré comme un moment d'information sur le dispositif de la formation. Les enseignants, bien qu'ils puissent participer de manière indirecte ou informelle à la suggestion de besoins, restent en dehors du processus technique et pédagogique de la conception et de la formalisation de l'action.
- Tout au long de la formation, bien que la réalité soit marquée par la diversité, les processus participatifs à caractère délibéré et intentionnel sont aussi relativement faibles et peu développés, sauf lorsqu'il s'agit de modalités de formation déterminées et avec certains formateurs.
- Cette étude révèle également que la phase finale de la formation devrait être un moment privilégié de l'évaluation (même si dans certains cas, une interaction régulatrice existe tout au long du processus de formation). Malgré ce constat, il s'avère que les résultats de la formation, notamment sous forme de rapports que les formés doivent rédiger, sont rendus après la fin de cette dernière, rendant impossible le dialogue entre les intervenants et les participants, ainsi que le feed-back du formateur.

Selon la modélisation que nous proposons, les résultats montrent que la PTE 1 tend à être dominante. Les pratiques d'évaluation, tout en relevant du paradigme objectiviste, se caractérisent par un ensemble de procédures non participatives, basées sur une logique de prise de décision bureaucratique, externe et non négociée. L'évaluateur a un statut externe et formel, mais dans le cas de la formation, il assume à la fois un statut interne et informel, sans conséquence en termes d'évaluation.

L'évaluation est donc essentiellement hétéronome. Cependant, on constate aussi un accroissement de l'autoévaluation située surtout à la fin du

processus de formation. Selon les participants, l'évaluation apparaît répartie entre la PTE 1 et la PTE 3, mais elle valorise davantage les fonctions consistant à contrôler, vérifier et mesurer. En tout cas, le contenu de l'évaluation constitue presque toujours le produit ou les *outputs* sous la forme de rapports écrits qui répondent plutôt à une exigence bureaucratique qu'à une nécessité pédagogique et formative. La PTE 2 apparaît comme une sorte de contrepoint uniquement rhétorique au niveau du discours, mais loin du monde empirique.

NOTE FINALE : LA RECHERCHE COMME UN PROCESSUS DE MODÉLISATION

Comme on s'en est aperçu, l'évaluation, bien que réalisée individuellement dans un contexte informel, n'est pas forcément égocentrique ; elle est en mesure de prendre en compte les partenaires et les organisations concernés ; elle a recours aux catégories de la culture dans laquelle s'insèrent ses protagonistes. La réflexion personnelle, dont elle découle et qu'elle exige, actualise un dialogue et une discussion sociale (et éducative) plus large, face à une expérience (personnelle) concrète qui aide à l'élaborer et avec laquelle elle évolue.

Les référentiels d'évaluation incorporent un ensemble d'axes d'analyse, ainsi que des critères qui permettent de décrire, de caractériser et de situer une pluralité de modèles de formation, dont la forme concrète qui est donnée à évaluer ne sera qu'une parmi de nombreuses possibilités. Mais les référentiels eux-mêmes, tout comme les modèles de référence, se construisent avec l'expérience et avec la réflexion que le processus même d'évaluation suscite. Ils ne sont pas figés et entièrement donnés d'emblée. Le travail dialectique d'évaluation cherche à opérer une synthèse qui pèse les avantages et les inconvénients de l'objet à évaluer en connaissance de cause, face à des modèles (idéaux types) qui ne sont pas parfaits, du fait qu'ils présentent, dans leur essence même, des points forts et des points faibles et que, par conséquent, ils sont sources d'ambivalence.

D'autre part, on a constaté que les dispositifs formels d'évaluation participative ne peuvent être considérés comme tels simplement parce qu'ils recueillent l'avis ou l'évaluation faite (individuellement) par les intéressés. En effet, cette forme d'évaluation peut se révéler d'un côté trop endiguée par les instruments de questionnement, et de l'autre, trop détachée des prises de décisions ou trop peu considérée par les autorités. Au lieu de privilégier la centration (subjectiviste) sur les personnes concernées, elle peut prendre un caractère purement formel et une nature essentiellement bureaucratique qui donnent seulement l'apparence et la sensation de participation. Les participants sont ainsi maintenus à la marge du pouvoir et de la construction soit de l'évaluation soit du curriculum de formation.

Si l'évaluation informelle menée individuellement par les participants peut se constituer de façon dialectique, comme le laisse entrevoir la première étude rapportée, la seconde révèle le poids de l'organisation sociale (politique) et institutionnelle de l'évaluation formelle. Celle-ci peut assujettir la participation individuelle et l'emprisonner dans des dispositifs, des systèmes et des procédures de nature objectiviste, autoritaire et bureaucratique (non dialogique et non démocratique).

Il semble donc qu'on ne peut se contenter des déclarations officielles ni des représentations induites par les schématisations théoriques à visée simplificatrice, même si l'on ne doit pas pour autant renoncer aux éclairages que ces dernières nous apportent. Il s'avère nécessaire et profitable de se consacrer à l'observation et à l'analyse des situations, dispositifs et pratiques concrètes dans leurs contextes d'existence, bref, de mettre en place une démarche de recherche pour les modéliser.

Nous estimons que les processus de modélisation de la participation à l'évaluation présentés dans ce chapitre font partie du processus générique de recherche qui passe par la confrontation (dialectique) d'un modèle théorique, hypothétique, avec les données provenant d'une situation réelle. Cette confrontation peut exiger l'adaptation du modèle, son ajustement et sa reformulation (Stufflebeam & Shinkfield, 2007). En effet, les modèles d'analyse de départ sont, on le sait, des simplifications utiles pour organiser la réflexion et penser l'action. Mais bien qu'indispensables, ils ne permettent pas de représenter avec précision des situations réelles de participation à l'évaluation, ni de décrire clairement les nuances et contours de situations spécifiques. Pour mieux apercevoir et se représenter le réel (en le modélisant), il faudrait peut-être, devant chaque cas, tester et ajuster les instruments d'observation, les grilles d'enregistrement, les cadres d'analyse et les règles d'interprétation, en bref, les modèles théoriques de référence (et/ou le référentiel d'évaluation). Les modèles issus de ces processus de modélisation se révéleront aussi importants, sans doute, pour aider à guider l'action. Une fois que de tels modèles se réfèrent à des pratiques, ils intègrent des référentiels et des modèles d'action qui peuvent être (dialectiquement) mobilisés dans la planification, la mise en marche, l'analyse, l'interprétation et la régulation de chaque action (du même type). Cette mobilisation intégrera, bien sûr, un processus de « re-modélisation » stratégique.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Bonniol, J.-J. & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles : De Boeck.
- De Ketele, J.-M. & Roegiers, X. (1999). *Metodologia da recolha de dados. Fundamentos dos métodos de observações, de questionários e de estudo de documentos*. Lisboa : Instituto Piaget.

- Fetterman, D. (2000). *Foundations of empowerment evaluation*. Thousand Oaks, CA : Sage.
- Figari, G. (1996). *Avaliar : Que referencial ?* Porto : Porto Editora.
- Figari, G. (2007). A Avaliação : história e perspectivas de uma dispersão epistemológica. In A. Estrela (Éd.), *Investigação em educação. Teorias e práticas (1960-2005)* (pp. 227-260). Lisboa : Educa.
- Greene, J. C. (2006). Evaluation, democracy, and social change. In I. F. Shaw, J. C. Greene & M. M. Mark (Éd.), *The Sage handbook of evaluation* (pp. 118-140). London : Sage.
- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. London : Sage.
- House, E. (1990). Methodology and justice. *New Directions For Program Evaluation*, 45(3), 23-36.
- House, E. R. & Howe, K. R. (2003). Deliberative democratic evaluation. In T. Kellaghan, D. L. Stufflebeam & L. A. Wingate (Éd.), *International handbook of educational evaluation* (pp. 79-100). Dordrecht : Kluwer Academic Publishers.
- Karlsson, O. & Conner, R. F. (2006). The relationship between evaluation and politics. In I. F. Shaw, J. C. Greene & M. M. Mark (Éd.), *The Sage handbook of evaluation* (pp. 225-242). London : Sage.
- Lesne, M. (1984). *Lire les pratiques de formation d'adultes*. Paris : Edilig.
- Lima, L. (1988). *Gestão das Escolas Secundárias. A participação dos alunos*. Lisboa : Livros Horizonte.
- Lima, L. (1998). *A escola como organização e a participação na organização escolar*. Braga : Universidade do Minho/Instituto de Educação.
- Machado, E. A. C. (2007). *Avaliação e participação. Um estudo sobre o papel dos actores na avaliação da formação contínua*. Thèse de doctorat, Instituto de Educação e Psicologia da Universidade do Minho.
- Mintzberg, H. (2010). *Estrutura e Dinâmica das Organizações* (4^e éd.). Lisboa : D. Quixote.
- Pique, B. (1990). *L'évaluation participative de la formation. Un outil du management*. Lyon : Chronique Sociale.
- Rodrigues, P. (1994). As três lógicas de avaliação de dispositivos educativos. In A. Estrela & P. Rodrigues (Éd.), *Para uma fundamentação da avaliação em educação* (pp. 93-120). Lisboa : Colibri.
- Rodrigues, P. (2001). L'entretien de recherche comme pratique d'évaluation. In G. Figari & M. Achouche (Éd.), *L'activité évaluative réinterrogée : regards scolaires et socioprofessionnels* (pp. 219-225). Bruxelles : De Boeck.
- Rodrigues, P. (2002). *Avaliação da formação pelos participantes em entrevista de investigação*. Lisboa : Fundação Calouste Gulbenkian/FCT-MCES.
- Rodrigues, P. (2004). Évaluation et formation des professeurs. Paradigmes d'évaluation (et de formation). In J. Ardoino & G. Berger (Éd.), *Former les enseignants et les éducateurs, une priorité pour l'enseignement supérieur [CD-ROM]*. Paris : AFIRSE/UNESCO/MEN.

- Rodrigues, P. (2006). Logiques d'évaluation en relation avec des conceptions élargies du monde et de la connaissance. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 193-201). Paris : L'Harmattan.
- Scriven, M. (2000). Evaluation Ideologies. In D. L. Stufflebeam, G. F. Madaus & T. Kellaghan (Éd.), *Evaluation models. Viewpoints on educational and human services evaluation* (pp. 249-278). Massachusetts : Kluwer Academic Publishers.
- Stake, R. E. & Schwandt, T. A. (2006). On discerning quality in evaluation. In I. F. Shaw, J. C. Greene & M. M. Mark (Éd.), *The Sage handbook of evaluation* (pp. 404-418). London : Sage.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco : John Wiley and Sons.

D'UNE ÉPISTÉMOLOGIE ABSENTE À UNE ÉPISTÉMOLOGIE EN DÉBAT

Enjeux épistémologiques, et donc aussi politiques, de la méta-évaluation

**Lise Demailly
Université de Lille 1**

Méta-évaluer est une opération réflexive qui s'origine d'un doute : et si l'évaluation que nous venons de réaliser était quelque part biaisée, peu pertinente, peu rigoureuse ? Et si les dispositifs d'évaluation que nous faisons fonctionner, par habitude ou obéissance, ne contribuaient pas au bien public autant que nous le pensons ? Mais qu'entendons-nous par méta-évaluation ? (Demailly, 2006). Imaginons qu'on a évalué, avec diverses méthodes, la qualité d'un programme de formation continue d'enseignants. Un doute a été exprimé quant à la pertinence des méthodes utilisées. On souhaite alors évaluer la qualité de cette évaluation et, d'une certaine façon, la modifier, l'améliorer. Ce faisant, on entre dans une démarche méta-évaluative. Méta-évaluer poursuit et approfondit le travail de réflexivité.

Le doute sur la qualité des évaluations est aujourd'hui fréquent de la part des acteurs sociaux qui font des méta-évaluations sauvages. La situation actuelle est donc paradoxale : d'un côté, jamais l'évaluation n'a été aussi légitime, autant perçue comme un remède à la crise des systèmes scolaires ou d'autres politiques publiques. De l'autre, les dispositifs d'évaluation sont critiqués, que ce soit quant à leurs outils (manque de rigueur scientifique des indicateurs), quant à leurs usages (lourdeur bureaucratique, illisibilité, coût), ou quant à leurs effets (déstabilisations stériles, stigmatisations liées aux palmarès, inapplicabilité des préconisations). Mais ce doute ne conduit pas toujours à réagir. Et paradoxalement, nous avons tendance à surtout méta-évaluer les évaluations qui nous sont imposées et, notamment, celles qui nous sont imposées de l'« extérieur ». Par exemple, PISA est l'objet

de nombreuses analyses critiques, sociologiques, épistémologiques. En revanche, faire la méta-évaluation d'une évaluation qui a semblé juste aux évaluateurs et qui n'a pas suscité colère ou résistance forte de la part des évalués, qui n'a pas « soulevé de vagues », est beaucoup plus rare. Cette situation est regrettable, car il serait judicieux de systématiquement juger après coup de la qualité et de l'utilité de toute évaluation. Pourtant, alors que la méta-évaluation est hautement instructive et souvent nécessaire pour l'enrichissement et l'achèvement du travail d'évaluation, il y a très peu de commandes officielles de méta-évaluation. Il faudra expliquer ce paradoxe : pourquoi, alors que la décision publique se légitime de plus en plus souvent par des connaissances et que l'évaluation est devenue une norme, se prive-t-on des connaissances évaluatives complémentaires produites par la méta-évaluation ? Nous verrons que ce paradoxe renvoie à la nature même de l'évaluation, et donc à celle de la méta-évaluation.

Dans ce texte, à partir de différents travaux que j'ai menés sur la transformation de l'action publique et de diverses expériences d'évaluation et de méta-évaluation, j'analyserai les enjeux épistémologiques – et donc également pratiques – de la méta-évaluation, en expliquant comment et pourquoi ils sont intimement mêlés et en montrant que la question épistémologique ne peut pas, de mon point de vue, être traitée de manière autonome.

LES OBJECTIFS COGNITIFS DES MÉTA-ÉVALUATIONS

Si le motif de l'analyse d'un dispositif d'évaluation est souvent un doute, quels sont les objectifs de cette opération de réflexivité ? À quelles questions faut-il répondre ?

On peut les synthétiser en trois points :

1) Premièrement, il s'agit d'émettre un jugement sur la nature et la qualité des objectifs qui ont été poursuivis. On souhaite en effet repérer les modèles et référentiels sous-jacents à l'action évaluée et à l'évaluation (modèles¹ cognitifs, éthiques, normes et type d'univers normatif...). Sont-ils pertinents ? Comportent-ils des malentendus, des imprécisions, des doubles langages, des présupposés non fondés ? Sont-ils cohérents entre eux ?

Il faut aussi clarifier les objectifs précis du dispositif d'évaluation. Par exemple, a-t-il été mis en place pour faire évoluer les pratiques, les savoirs et les représentations ou bien pour recueillir des données qui vont informer un responsable hiérarchique ? En vue d'une prise de décision

1. J'entends par modèles des guides pour l'agir. Les mondes sociaux se caractérisent par la coexistence concurrentielle de nombreux modèles, implicites ou explicites, tacites ou formalisés, experts ou profanes.

ultérieure ? En vue de sanctions ? Pour la gestion ? S'agissait-il d'évaluer des personnes ou des dispositifs ? Ces objectifs étaient-ils conscients ? Clairs ? Pertinents ? Est-ce que tous les acteurs concernés se représentaient les objectifs du dispositif d'évaluation de la même façon ? N'y avait-il pas trop d'objectifs, et notamment des objectifs hétérogènes impliquant des paradigmes inconciliables ?²

**EXEMPLE : LE CAS DES AUDITS D'ÉTABLISSEMENTS
SCOLAIRES DE L'ACADÉMIE DE LILLE**

Dans les audits lillois d'établissements scolaires effectués entre 1991 et 1995 dans les 519 établissements du second degré, à la demande du recteur de l'académie de Lille (Demailly, 1999), certains organisateurs de l'évaluation voyaient une évaluation *sommative* destinée à dresser un état des lieux et à *orienter la décision* rectorale. D'autres voyaient surtout une évaluation *formative* en direction de chacun des établissements audités pour dynamiser les projets d'établissement. Les deux objectifs ont coexisté de manière confuse. Or, les méthodologies impliquées auraient été différentes pour chacun de ces objectifs s'ils avaient été clairement formulés, et elles étaient difficilement conciliables au sein d'une même opération. Le second a donc pris finalement le pas sur le premier.

Au niveau des référentiels, l'audit introduisait subrepticement, sans délibération explicite, par le biais de la notion d'« image de l'établissement » qui était un des items de l'audit, une mise en compétition des établissements, norme ne correspondant pas au référentiel de solidarité de service public qui réunissait alors une majorité d'établissements (à l'exception des plus bourgeois). Cette ambiguïté (ambiguïté plutôt que manipulation) s'est révélée ultérieurement source de tensions.

Les évaluations concrètes pratiquent souvent le mélange des genres, la « confusion des langues ». L'analyse des objectifs a donc une finalité de *clarification pratique* – éthique même pourrait-on dire – et méthodologique en même temps. Elle produit aussi des connaissances scientifiques sur la manière dont les acteurs sociaux négocient leurs objectifs et leurs modèles, les marchandent, les masquent intentionnellement, ou alors ne les perçoivent même pas consciemment.

2) Ensuite, et ceci constitue le deuxième axe de travail, méta-évaluer implique d'émettre un jugement sur la cohérence épistémologique et méthodologique du dispositif d'évaluation étudié, la qualité de la démarche

2. Pour la notion de paradigmes inconciliables, voir la contribution d'Allal dans le présent ouvrage.

et des outils et leur capacité à produire des données et des représentations exactes de la réalité, ainsi que sur le repérage des biais et des limites.

3) Le troisième objet d'analyse d'une méta-évaluation concerne l'ensemble des effets du dispositif d'évaluation. Autrement dit, il faut répondre aux questions suivantes :

- Les connaissances attendues ont-elles été produites ?
- Les effets pratiques attendus ont-ils eu lieu ? Si des décisions étaient attendues, ont-elles pu être prises ?
- D'autres effets, positifs ou négatifs, non prévus, pervers, sont-ils observables ?

Ce troisième volet de l'analyse produit une *critique politique* des dispositifs d'évaluation. Elle montre éventuellement que le dispositif n'a servi à rien (les rapports présentant leurs résultats dorment dans les placards, le même rapport est recommencé plusieurs fois à quelques années d'intervalle, sans plus de succès pratique³, ou sert essentiellement à ceux qui en avaient le moins besoin⁴). L'étude peut également mettre en évidence les raisons des résistances à un dispositif ou ses effets pervers. Elle peut aussi souligner l'intérêt de certains résultats (concernant des phénomènes insoupçonnés, des injustices ou des préjugés, révélés par l'évaluation) ou des effets de dynamisation heureuse d'une ampleur initialement imprévisible. Dans la recherche européenne *Knowledge and Policy*⁵, nos collègues belges ont pu montrer à quel point les mauvais résultats de la Communauté française de Belgique à PISA avaient entraîné un mouvement de réforme de l'enseignement belge (Mangez & Cattonar, 2011), une mobilisation large.

Ce troisième objet des méta-évaluations, à savoir les effets de l'évaluation, est essentiel. Car l'étude des effets a une valeur pratique : l'évaluation a-t-elle servi à quelque chose et à quoi ? Mais cette critique politique a en même temps une valeur cognitive : elle produit des données sur la manière et les processus selon lesquels les acteurs sociaux ou les décideurs politiques s'emparent ou non d'un dispositif d'évaluation et de ses résultats.

Dans leurs trois dimensions, les méta-évaluations ont donc en principe une valeur indissolublement cognitive et pratique, elles permettent d'aller

3. Dans le domaine de la santé mentale, il y a eu depuis 15 ans une dizaine de rapports d'évaluation adressés aux ministres de la Santé, le dernier en date étant celui de la Cour des comptes, pour faire les mêmes constats et émettre les mêmes recommandations, sans que rien ne change.

4. Le dispositif d'audit des établissements scolaires de l'académie de Lille qui a été déployé entre 1991 et 1995 a ainsi servi à dynamiser essentiellement les établissements déjà dynamiques (Demailly, 1999).

5. Contrat de recherche européen sur les rapports entre connaissances (dont l'évaluation) et politiques, dans le domaine de l'éducation et de la santé mentale : <http://www.knowandpol.eu/>

plus loin que les dispositifs d'évaluation, de les rectifier, de les améliorer, de poursuivre la dynamique de réflexivité. Les objectifs cognitifs et les objectifs pratiques et politiques sont souvent complètement entremêlés dans les méta-évaluations concrètes (sauf peut-être dans le cas de la sociologie de l'évaluation pour laquelle les objectifs sont principalement scientifiques).

LA MÉTHODOLOGIE DES MÉTA-ÉVALUATIONS

Quelques principes méthodologiques, qui ont bien sûr une implication épistémologique, me semblent devoir être respectés pour une démarche rigoureuse et fournir les réponses à la série de questions listées à la section précédente. Méta-évaluer suppose en principe une démarche d'étude scientifique, utilisant des méthodes d'investigation, d'enquête et de mesure classiques. Mais la méta-évaluation proprement dite ne s'achève qu'avec le jugement de valeur qui peut être porté sur l'objet, en l'occurrence sur le dispositif d'évaluation méta-évalué. Ce jugement de valeur est énoncé par une instance décisionnelle commanditaire, après l'étude, en fonction d'elle, ou par l'évaluateur lui-même. La démarche d'ensemble ne saurait donc être neutre. La méta-évaluation a toujours, comme l'évaluation, une finalité pratique et normative d'exercice du jugement ; ce n'est pas de la « recherche pour la recherche », même si elle s'appuie sur une recherche.

Soutenir une distance critique

Il est souhaitable que l'opérateur de l'étude soit indépendant, universitaire de préférence, capable d'objectiver et de prendre une distance critique avec les pratiques d'évaluation étudiées et, dans l'idéal, qu'il ne poursuive aucun enjeu personnel pratique dans le champ étudié, ce qui n'est pas toujours le cas des méta-évaluations concrètes.

Cette distance n'exclut pas d'utiliser éventuellement, à certains moments de l'étude, une méthodologie participative, ni non plus – cela paraît même tout à fait souhaitable – de recueillir l'avis des différents participants au processus d'évaluation : commanditaire, maître d'œuvre, personnes collectant les données, personnes évaluées directement ou indirectement peuvent ainsi être invités à exprimer leur avis sur l'évaluation, ou leur vécu de celle-ci. La distance épistémologique est à conquérir et soutenir, car le méta-évaluateur a souvent déjà des affinités avec certains acteurs du champ, s'il lui est familier, voire des enjeux personnels pratiques s'il enchaîne des contrats avec les acteurs d'un champ dont il est alors devenu un expert. Il faut donc que, dans ce cas, il soit capable d'objectiver sa propre position, d'analyser de façon réflexive ses affinités, ses présupposés, ses intérêts, d'explicitier en quoi ses propres valeurs orientent son travail intellectuel.

Il n'est pas question de rechercher une neutralité impossible, mais de se montrer clair par rapport à des engagements conscients, honnêtes et lucides par rapport aux effets de position.

Saisir le dispositif d'évaluation comme une pratique sociale complexe

Un dispositif d'évaluation met en jeu des intérêts, des valeurs fort diverses (valeurs professionnelles, valeurs politiques, valeurs de vérité...), des idées⁶, des connaissances, un grand nombre d'acteurs et des cultures⁷. Il sera souvent pertinent de pratiquer d'abord une analyse stratégique, au sens que Crozier et Friedberg (1977) donnent à ce terme, pour repérer les jeux d'intérêt et les rapports de force ; puis de mener des analyses complémentaires pour les éléments qui ne seraient pas réductibles à une simple logique d'intérêts ou de pouvoir (pour plus de détails sur ce point du « stratéguisme méthodologique », voir Demailly, 2005), et qui concernent les représentations des acteurs et leurs valeurs.

Le processus d'évaluation apparaît en général comme un processus négocié, qu'il s'agisse de négociation clinique ou de négociation politique⁸. Cette négociation peut être, dans certains cas, relativement égalitaire : on peut alors analyser les processus d'apprentissage qui accompagnent le travail d'évaluer ou celui de recevoir une évaluation, mettre en évidence le rôle fondamental de la confiance entre évaluateurs et évalués pour que l'évaluation puisse devenir la coconstruction d'un consensus sur ce qu'est la « qualité » (Demailly, 2003) et produire des modifications effectives de pratique (pédagogique, organisationnelle...), de posture, de vision du monde professionnel ou des coacteurs (les élèves, les partenaires, l'administration, etc.). Mais l'évaluation peut aussi être le lieu de rapports de force brutaux, qui se dissimulent derrière l'apparente objectivité de la mesure.

6. Pour ce triptyque intérêts, valeurs, idées, dans l'analyse des politiques publiques, voir Palier et Surel (2005).

7. Par exemple, dans le domaine psychiatrique, la méta-évaluation de la plupart des échelles destinées à mesurer la paranoïa non seulement fait apparaître une fidélité inter-juges qui peut sembler insuffisante et montre que l'outil d'évaluation est imparfait, mais surtout sa mise en jeu dans un pays comme Madagascar révèle un taux de paranoïa dans la population 10 fois supérieur à ceux des pays développés. Cela a montré que l'outil d'évaluation, qui se prétendait a-culturel et universel, est en fait marqué par un biais culturel. Si les méta-évaluateurs ne s'étaient pas intéressés à ce type de données, apparemment éloigné de la question initiale, ils n'auraient pu repérer et signaler cette faiblesse qui demandait correction. (Observation de terrain réalisée par l'auteure, Congrès de psychiatrie, Tananarive, juin 2004.)

8. Voir les contributions d'Allal et de Merle dans le présent ouvrage.

Prendre le dispositif d'évaluation comme un processus temporel

Il est souhaitable d'étudier le dispositif d'évaluation tout au long de son déroulement, depuis la formation de l'idée qu'il serait utile d'évaluer une action, la mise en forme de la commande et l'éventuel appel d'offres, la négociation ou la non-négociation avec les acteurs concernés par l'évaluation des outils d'évaluation, la mise en œuvre de la méthodologie, la mise en forme des résultats, que ceux-ci soient ponctuels ou récurrents, jusqu'aux effets dans l'après-coup, bien au-delà de la seule mise en circulation des résultats⁹. Je défends donc, sur le plan épistémologique, une approche large de l'objet d'évaluation dans la méta-évaluation, même s'il faut ensuite répondre à des questions précises comme celle de la pertinence, de la validité ou de la fidélité de tel outil pour mesurer tel référent. Le méta-évaluateur peut être amené à faire intervenir des phénomènes ou à étudier des aspects qui peuvent paraître très éloignés de l'objet initial.

Repérer les problèmes de positionnement de la méta-évaluation

Il reste que, comme les évaluations, les méta-évaluations sont tendues entre deux pôles en ce qui concerne leur commande, et donc leur cadre symbolique. Soit, et ceci correspond au premier pôle, il s'agit d'une auto-commande de chercheurs ou d'une commande scientifique. On observe que la sociologie critique de l'évaluation, que j'avais appelée de mes vœux en 1995, se développe effectivement, comme en témoignent la publication de plusieurs numéros de revues ou le développement de colloques dont l'objet est l'analyse sociologique de l'évaluation¹⁰. Si la position symbolique de la méta-évaluation est clairement celle d'une recherche, la qualité épistémologique peut en être excellente, la production de connaissances scientifiques incontestable. Dans ce cas, l'utilité pratique n'est pas automatique, elle ne peut venir que de surcroît et de manière en partie aléatoire si le travail de méta-évaluation rencontre des acteurs sociaux qui se saisissent de cette possibilité de lucidité accrue. Cela requiert un certain nombre de conditions, des hasards heureux et des volontés politiques.

9. Dans le présent ouvrage, la contribution de Lafontaine et Monseur sur PISA est un exemple de mise en œuvre des points 2 et 3 du programme de travail exposé ici.

10. C'est le sujet essentiel des XIII^{es} Journées internationales de sociologie du travail, « Mesures et démesures du travail », Bruxelles, 25-27 janvier 2012 : <http://metices.ulb.ac.be/spip.php?rubrique46>

Soit, et ceci constitue le second pôle, il s'agit d'une commande institutionnelle et, comme l'évaluation, la méta-évaluation fait alors partie de l'institution, avec la duplicité foncière que cela implique. Les chercheurs ne peuvent alors pas toujours mettre en application les principes épistémologiques d'une méta-évaluation rigoureuse. Une méta-évaluation commandée par l'institution peut se trouver contrainte de ne produire que des savoirs praxéologiques améliorés, car elle ne peut se déployer dans ses différentes dimensions. Assez souvent, l'institution ne veut pas s'intéresser aux effets pratiques de l'évaluation, ni mesurer sa productivité. Très occupées à légitimer l'évaluation, les institutions ne veulent en effet pas prendre le risque que des chercheurs révèlent qu'une évaluation coûteuse n'a pas servi à grand-chose.

En cas de méta-évaluation encadrée institutionnellement – il en va de même pour l'évaluation – si le travail produit des connaissances, celles-ci ne sont pas vraiment de nature scientifique, elles sont plutôt de type pédagogique, technique ou communicationnel.

EXEMPLE : LE CAS D'UN SÉMINAIRE *KNOW AND POL* EN 2011

Lors d'une réunion avec les utilisateurs finaux, organisée dans le contrat européen *Know and Pol* (cf. note 5 *supra*), les cadres supérieurs d'un grand organisme international ont vivement réagi à une méta-évaluation qui montrait que les évaluations qu'ils pratiquaient et publiaient n'étaient pas rigoureuses épistémologiquement. Ils affirmèrent qu'ils le savaient, que leurs évaluations (statistiques et internationales) avaient effectivement un but essentiellement pédagogique, et que cela n'avait aucun intérêt pour eux qu'on leur dise que ce n'était pas scientifique. Ils assumèrent ainsi pleinement le caractère praxéologique des connaissances qu'eux-mêmes produisaient. Ils appelèrent le chercheur méta-évaluateur à en faire autant, à ne pas faire de la « recherche pour la recherche », et à produire des connaissances utiles. (Observation réalisée par l'auteur d'un séminaire *Know and Pol*, 2011.)

J'appelle connaissances « praxéologiques » des connaissances produites par des « technologies du social » (Demailly, 2000). Ces technologies du social utilisent largement des outils qui sont par ailleurs insérés dans des démarches scientifiques, comme la statistique, ainsi que des procédures de formalisation, mais le but est prioritairement pratique. L'expert ne peut se soutenir dans l'autonomie du champ scientifique (valeurs de vérité) et il accepte que l'institution lui fixe ses objectifs (valeurs d'utilité), voire ses concepts.

Paradoxalement, il n'est pas sûr que les connaissances produites dans les méta-évaluations institutionnellement bridées soient réellement plus utiles que les évaluations étudiées qu'elles contribuent à répéter et légitimer telles quelles. Elles le sont sans doute davantage pour les commanditaires et les responsables hiérarchiques que pour les professionnels de base, encore que les commanditaires puissent se dire « déçus » par les chercheurs quand ceux-ci ne leur fournissent pas un *package* d'idées directement applicables. En somme, ce n'est pas parce qu'une méta-évaluation vise l'utilité qu'elle est utile. On peut même affirmer qu'elle a moins de chances de l'être que dans les cas où le chercheur a les moyens de mettre cette préoccupation de côté et de satisfaire essentiellement sa curiosité. Bref, une méta-évaluation est plus utile quand elle ne sombre pas dans l'insignifiance mais introduit une rupture dans les manières de penser, ce qu'une approche scientifique est davantage en mesure de produire. Le souci de viser une certaine utilité, d'accroître la lucidité des professionnels de base et des acteurs de terrains, à partir de leurs propres questions, tout en gardant une distance critique vis-à-vis des institutions, peut en somme constituer un principe épistémologique correct. Il revient finalement à tenir et soutenir une certaine éthique du métier de chercheur.

QUELQUES RÉSULTATS GLOBAUX DES MÉTA-ÉVALUATIONS

Je voudrais maintenant présenter quelques résultats transversaux des méta-évaluations, qui montrent leur intérêt scientifique et donc aussi leur utilité sociale.

Critique épistémologique de dispositifs particuliers

Les méta-évaluations montrent souvent que les outils d'évaluation ne sont pas « scientifiques » ou que la présentation des résultats n'est rigoureuse qu'assortie de nombreuses spécifications. Cela vaut aussi bien pour les méthodologies qualitatives que quantitatives.

Alors qu'on pourrait penser que les techniques qualitatives sont un moyen privilégié pour comprendre en « profondeur » le fonctionnement d'un établissement scolaire, l'expérience et l'observation du dispositif d'évaluation en acte montrent concrètement la difficulté d'une telle démarche dès lors que : i) celle-ci est mise en œuvre par des agents qui n'y sont pas spécialement et solidement formés ; ii) qu'elle est effectuée sur une base de questionnements très standardisés dans leur forme, mais pas dans leur passation ; et iii) que les agents l'improvisent soit par souci de prudence

avec force « garde-fous », soit en forçant des réponses sans s'en rendre compte. Les techniques d'investigation qualitative ne supportent guère le flou méthodologique.

**EXEMPLE : LE CAS DES AUDITS
D'ÉTABLISSEMENTS SCOLAIRES LILLOIS**

La méta-évaluation du dispositif d'audit des établissements scolaires lillois (Demaillly, 1999) montre qu'il s'agit d'une évaluation non scientifique, partiellement négociée, dont la finalité devient en cours de processus de plus en plus formative, ne produisant pas de connaissances objectives pour éclairer les décisions du recteur, mais effectuée à des fins de dynamisation des professionnels de base, objectifs atteints partiellement (sur les établissements déjà dynamiques et sur les évaluateurs eux-mêmes).

Le risque de « dérapage » est sans cesse évoqué dans les réunions de synthèse et dans les séminaires de formation à l'usage de ceux qui vont réaliser les audits : il est recommandé d'éviter dans les questions de provoquer « lieux communs », « rancunes », « catalogue de revendications syndicales », « affrontements stériles ». Du coup, les évaluateurs s'autocensurent dans le questionnement et le recueil des données.

Évitant le réductionnisme technocratique, une telle approche globale et humaniste manque de vertu objectivante, autrement dit ne produit pas les effets de rupture et de remise en cause que permettent des instruments quantitatifs centrés sur certains résultats scolaires des élèves.

La méta-évaluation de dispositifs basée sur des mesures quantitatives révèle souvent des biais méthodologiques et épistémologiques ; cela est bien connu et je ne m'y attarderai pas.

Enfin, l'analyse des évaluations permet de voir les limites respectives et croisées des référentiels de l'obligation de moyens et de l'obligation de résultats quand ils passent par des indicateurs chiffrés. Dans le premier cas, le respect des normes en termes de nombre d'actes ou d'utilisation de telles procédures ne garantit rien quant à la qualité des résultats, qui s'invisibilisent. Dans l'autre, c'est le travail lui-même et les habiletés qui y sont déployées qui sont invisibilisés. Chaque évaluation met donc dans l'ombre un non-évalué qui est parfois très important et produit des biais de connaissance et de reconnaissance.

**EXEMPLE : LE CAS DE L'ÉVALUATION
D'UNE ASSOCIATION D'INSERTION SOCIALE**

« Si le financeur constate, année après année, que l'insertion n'est pas au rendez-vous [...], quelles conclusions en tirera-t-il ? Le risque est grand qu'il incrimine la structure taxée d'incompétence et ses professionnels jugés peu efficaces. Alors même qu'une évaluation qualitative de l'insertion sociale permettrait de créditer l'association évaluée et ses professionnels d'une véritable habilité dans le champ de compétences qui est le leur [...]. L'évaluation quantitative du devenir social professionnel *ex post* fait totalement abstraction de l'action au quotidien des intervenants. Le but en fait n'était pas d'identifier les méthodes de travail les plus performantes par rapport à des populations très défavorisées, mais de réduire les coûts. » (Frétigné, 2012, p. 10)

Globalement les défauts épistémologiques des évaluations apparemment scientifiques proviennent :

- De l'existence de « trous » (ce qui est mis à l'ombre) aboutissant à des biais ou à une perte de sens ;
- De la confusion des genres, de l'hypocrisie ou de la juxtaposition d'objectifs qui auraient dû être, d'une part, explicités, d'autre part, traités séparément (à condition d'être avouables, ce qui n'est pas toujours le cas).

Efficacité des évaluations

Un acquis des méta-évaluations est la mise au jour des conditions permettant à une évaluation d'être efficace en termes de transformation positive des pratiques, ce qui implique une certaine qualité des savoirs produits. Ce sont, d'une part, la pluralité des modes d'approche, d'autre part, le caractère participatif, collaboratif de l'évaluation.

LA PLURALITÉ DES MODES D'APPROCHE

Il faut d'abord qu'une évaluation soit pluraliste et large. En effet, si elle ne met en jeu que les moyens, procédures, méthodes employées, elle censure la question des résultats finaux. Si elle est focalisée sur les résultats, elle met souvent en cause l'expertise et l'éthique des professionnels. Au niveau des effets sociaux, elle produit des effets pervers, les professionnels ayant les yeux sur le compteur et non plus sur la route. Une pratique gouvernée par les indicateurs est une pratique dont la qualité se dégrade. Elle se prête à la tricherie et la manipulation par rapport aux chiffres. Ces mécanismes, aujourd'hui, concernent l'ensemble des métiers.

EXEMPLES : DE LA FINANCE À LA SANTÉ

Les « challenges » à la Banque Postale conduisent les conseillers financiers à placer des produits dont parfois leurs clients n'ont pas besoin.

Quant aux médecins généralistes, ils sont depuis peu obligés à un certain nombre d'actes préventifs pour toucher une prime s'ils atteignent les chiffres prévus ; mais on n'évalue pas si la santé ou les soins de la population dont ils s'occupent se sont améliorés grâce à ces mesures.

Il faut donc que les dispositifs d'évaluation saisissent la réalité dans sa complexité, sans la simplifier, même s'ils opèrent obligatoirement des réductions. Il y a un juste équilibre à trouver entre élaboration d'outils de mesure et respect de la complexité du travail (que ce soit celui des professionnels ou celui des élèves).

LE CARACTÈRE PARTICIPATIF DE L'ÉVALUATION

La participation est nécessaire à la qualité des données recueillies, à la qualité de leur interprétation, ainsi qu'à la saisie de données qualitatives dont je pense qu'elles apparaîtront de plus en plus clairement comme le complément indispensable de la mesure chiffrée. Elle est aussi nécessaire au bon accueil de l'évaluation. Elle s'inscrit alors dans une « régulation autonome ». Elle est, enfin, nécessaire à l'appropriation de l'évaluation, au développement de ses effets sociaux et, notamment, de ce que j'appellerais volontiers non plus « culture de l'évaluation », mais « compétences d'autoévaluation ». Il m'a été donné d'observer la réussite d'un dispositif d'évaluation de contrats de zones d'éducation prioritaire qui a respecté ce principe (Demailly, 2003) et, à l'opposé, de multiples contre-exemples.

Les méta-évaluations permettent de distinguer les caractéristiques des évaluations qui sont efficaces, au sens où elles ne se contentent pas d'apporter des connaissances ou de seulement permettre à l'administration de rationaliser sa gestion, ni ne restent lettre morte dans des logiques où les rapports d'évaluation se répètent et s'accumulent, mais qu'elles déclenchent des mobilisations concrètes pour comprendre un problème, remédier à un dysfonctionnement. Les résistances à l'évaluation ont plusieurs sources, qui concernent à la fois les intérêts des personnels des systèmes éducatifs, leurs valeurs, les ressources cognitives qu'ils estiment utiles, et parfois même leur bon sens. Il semble bien que les dimensions décisives soient celle de la participation d'une part, celle des valeurs partagées de l'autre, car intéresser des professionnels à l'évaluation de leur propre pratique ne peut se faire que sur la base d'une réelle confiance politique. Il est également nécessaire que

les évaluateurs acceptent d'écouter et d'apprendre des évalués, y compris sur la fabrication des outils d'évaluation et sur les méthodes de dialogue autour de ces évaluations.

La nature de l'évaluation

L'évaluation n'est ni une pratique stratégique ordinaire, ni un dispositif scientifique, c'est une pratique institutionnelle, une institution, et, en ce sens, elle réalise une connexion tout à fait spécifique et originale de la production de connaissance et de la production de pouvoir. Ce n'est pas une pratique de connaissance pure et désintéressée, c'est une pratique de connaissance stratégique, politique. Elle est prise aujourd'hui dans les enjeux des modernisations des appareils d'État et de la construction européenne, dans une période de forte poussée des modes de réflexion néo-libéraux sur la gouvernance des sociétés et leur mise en compétition. Elle est prise aussi, plus concrètement, dans les enjeux liés à la redéfinition des priorités, des finalités, des valeurs centrales des politiques éducatives. Bref, elle a un sens politique négocié ou conflictuel et ne peut être considérée globalement comme scientifique, même si elle contient des éléments d'étude scientifique, qui peuvent d'ailleurs être réutilisés ailleurs, dans d'autres synthèses, dans le champ scientifique.

En revanche, l'évaluation n'est pas une pratique sociale seulement stratégique, elle est guidée par des valeurs de vérité, de connaissance, de production de savoirs sur le réel. L'idée même d'évaluation telle qu'elle s'est développée à la fin du xx^e siècle appartient à la série des technologies rationnelles de l'action (Demailly, 2000), qui permettent d'agir de manière plus rationnelle précisément grâce à une production de connaissances sur les réalités humaines et sociales.

L'idée d'évaluation est celle d'une connexion entre savoir et pouvoir *d'agir pertinent*. On voit qu'elle implique une représentation du monde très différente de celle où le pouvoir d'agir pertinent (et légitime) était reconnu comme relevant de l'inspiration par exemple, ou de la grâce, de la naissance, de la révélation, ou, avec l'avènement de la légitimité bureaucratique, du respect du règlement ou de l'ordre. L'évaluation est donc une forme institutionnelle, la forme institutionnelle de la modernité, quand celle-ci s'engage dans la rationalisation, la formalisation, la standardisation de l'action et de la relation (Demailly, 2008). En ce sens, elle est aujourd'hui un dispositif central de *gouvernementalité*, de gouvernement des conduites (Foucault, 2004).

Réalisant une mise sous observation du travail, une mise sous observation de la réalité sociale, l'évaluation a intérêt à ce que la connaissance qu'elle produit soit la plus exacte possible, en fonction des objectifs

pratiques qu'elle se fixe, d'où un certain souci de rigueur. Mais en même temps, et à cause de sa nature institutionnelle, les dispositifs d'évaluation sont toujours performatifs, et comportent toujours des « trous », des zones de méconnaissance, des zones de censure. L'observation de l'évaluation des politiques publiques montre que celle-ci présente des lacunes significatives, que des connaissances qui seraient importantes pour l'action sont manquantes. La méconnaissance de ces « trous » produit des artefacts, des évaluations dépourvues de sens réel, des mesures creuses et parfois insensées. Les « trous » bien souvent se comprennent par rapport à la nature institutionnelle de l'évaluation, aux valeurs et fonctions dominantes qu'elle sert, comme le montre clairement l'exemple suivant, emprunté au domaine de la police : les indicateurs utilisés concernent uniquement le nombre d'« affaires » traitées et résolues, mais pas la mesure du temps d'attente au téléphone pour un usager ni le sondage auprès des citoyens sur leur expérience du travail policier (Mouhanna, 2011). Autrement dit, la nature institutionnelle de l'évaluation fait que les dispositifs d'évaluation visent principalement, aujourd'hui, la mise sous observation du travail et le contrôle des professionnels, en privilégiant la commodité d'indicateurs simples et automatisés.

La deuxième tendance lourde est de faire de plus en plus fonctionner un référentiel marchand, qui met en compétition les individus les uns avec les autres et les organisations entre elles. En revanche, les évaluations qui seraient vraiment utiles aux professionnels (ou celles qui pourraient obliger les supérieurs hiérarchiques et politiques à se mettre en cause) ne sont pas faites, ou faites de telle façon qu'elles découragent les acteurs de se les approprier.

CONCLUSION : SCIENTIFICITÉ DES MÉTA-ÉVALUATIONS ?

Je voudrais pour terminer répondre à une question centrale de cet ouvrage : les méta-évaluations et les évaluations peuvent-elles être scientifiques et à quelles conditions ? Ces conclusions valent aussi de manière plus générale pour l'évaluation, dont les méta-évaluations ne sont qu'un cas particulier.

Parfois, la réponse est positive. Les méta-évaluations et les évaluations peuvent produire des connaissances scientifiques, au prix d'un dégagement par rapport à la nature institutionnelle de l'évaluation, dégagement qui implique pour le chercheur d'assumer des tensions avec ses commanditaires, de reformuler la demande, d'échapper au cadre cognitif qu'on lui propose, voire qu'on essaie de lui imposer. Car l'évaluation, à l'époque contemporaine, sert surtout à légitimer l'ordre établi et les valeurs dominantes et à rappeler tout un chacun à la subordination. Même quand elle paraît

critiquer l'existant, elle cherche surtout à contrôler les professionnels et à limiter leur autonomie, dans le sens des fonctionnements dominants qui, eux, ne sont pas évalués ni mesurés. Par exemple, on ne mesure jamais le coût économique exorbitant du travail d'évaluation, et on évalue rarement les politiques publiques. La commande officielle de méta-évaluation, quand elle existe, n'est souvent que le signe d'une insatisfaction institutionnelle par rapport à une évaluation : les résultats ne sont pas assez ou trop sévères, les chercheurs n'en ont fait qu'à leur tête, il y a eu trop de critiques sur l'évaluation. Elle invite rarement le chercheur à exercer son esprit critique en toute liberté. Une méta-évaluation strictement praxéologique et contrôlée par les institutions, mais bien faite, peut certes être utile à des commanditaires en mal de formules-chocs, de « visions » ou de raffinement de leurs outils de gouvernement, ou/et elle peut aller dormir dans un placard, à côté du rapport d'évaluation qu'elle a étudié, mais elle sert mal la réflexivité démocratique, c'est-à-dire la capacité d'une société à réfléchir, dans le cadre de débats démocratiques, à ses choix de finalités et à ses arbitrages de moyens.

En revanche, quand elle est rigoureuse et libre, l'analyse des évaluations peut avoir une utilité politique : permettre aux acteurs sociaux de mieux résister à l'emprise institutionnelle de la plupart des dispositifs d'évaluation, en en critiquant, le cas échéant, le manque de rigueur scientifique, les effets sociaux parfois pervers ; identifier les évaluations utiles tant dans leur contenu que dans leur forme ; relancer la réflexion. De même, elle permet aux acteurs professionnels de base, aux coacteurs et partenaires, de participer à la construction des indicateurs, d'introduire leur parole et leur point de vue dans les dispositifs d'évaluation, de faire progresser une évaluation pluraliste et partagée.

La méta-évaluation peut également être productrice de connaissances scientifiques dans les domaines d'action concrète concernés, y compris ceux qui ont tendance à rester dans l'ombre, je pense par exemple aux questions d'inégalités ou de discriminations. Elle peut aussi plus largement, comme j'ai tenté de le montrer, faire la lumière sur le « gouvernement politique par les instruments » (Lascoumes & Le Galès, 2005), sur les processus sociaux de négociation de l'action publique, sur la construction conflictuelle des représentations et des modèles légitimes.

Les deux aspects peuvent être liés... ou non. Il n'est pas toujours certain qu'une méta-évaluation riche sur le plan des connaissances scientifiques rencontre les acteurs sociaux auxquels elle pourrait être utile en leur apportant un surcroît de lucidité. Il n'est pas certain non plus qu'elle rencontre la volonté politique. Les logiques de la connaissance et celles de l'action subissent quelques décalages, et ce serait pur idéalisme que de croire que des connaissances vraies entraînent inmanquablement des actions justes.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Crozier, M. & Friedberg, E. (1977). *L'acteur et le système*. Paris : Seuil.
- Demailly, L. (1999). L'évaluation et l'auto-évaluation des établissements : un enjeu collectif. Le cas des audits d'établissements scolaires de l'académie de Lille. *Politique et management public*, 17(1), 37-58.
- Demailly, L. (2000). Le mode d'existence des techniques du social. *Cahiers internationaux de sociologie*, 153, 103-124.
- Demailly, L. (2003). L'évaluation comme apprentissage et négociation. *Revue française de pédagogie*, 142, 115-129.
- Demailly, L. (2005). Pour une représentation politique de l'acteur au travail. *Cahiers lillois d'économie et de sociologie*, 45, 153-174.
- Demailly, L. (2006). La nécessaire méta-évaluation. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 203-211). Paris : L'Harmattan.
- Demailly, L. (2008). *Politiques de la relation. Sociologie des métiers et du travail de relation*. Lille : Presses Universitaires du Septentrion.
- Foucault, M. (2004). *Sécurité, territoire, population. Cours au Collège de France, 1977-1978*. Paris : Seuil.
- Frétiigné, C. (2012, janvier). *Quand « évaluation de dispositifs d'insertion » vaut « évaluation du travail de chargés d'insertion »*. Une mesure du travail en creux. Communication présentée aux XIII^{es} Journées internationales de sociologie du travail (JIST), Bruxelles.
- Lascoumes, P. & Le Galès, P. (Éd.). (2005). *Gouverner par les instruments*. Paris : Presses de Sciences Po.
- Mangez, E. & Cattonar, B. (2011). La place de PISA dans le rapport entre la société civile et les acteurs de l'enseignement en Communauté française de Belgique. *Sciences de la société*, 79, 29-41.
- Mouhanna, C. (2011). *La police contre les citoyens ?* Paris : Champ social.
- Palier, B. & Surel, Y. (2005). Les « trois I » et l'analyse de l'État en action. *Revue française de science politique*, 55(1), 7-32.

Les exigences inconciliables des activités évaluatives et leurs paradigmes de référence

Linda Allal
Université de Genève

Il existe deux grands champs de l'évaluation en éducation. Le premier est l'évaluation des caractéristiques (connaissances, compétences, démarches, attitudes...) de personnes en vue d'actions ou de prises de décisions qui les concernent individuellement. Ce champ comprend l'évaluation des apprentissages des élèves et étudiants par un formateur, ainsi que l'évaluation des compétences des enseignants ou d'autres professionnels par un directeur d'école, conseiller pédagogique, chef de service, etc. Le second champ est celui de l'évaluation des dispositifs de formation, expression qui englobera ici l'évaluation des programmes, méthodes et outils didactiques, projets d'intervention, innovations, établissements, systèmes éducatifs. Dans ce second champ, l'évaluation est effectuée par des agents (le plus souvent une équipe composée de chercheurs et/ou administrateurs), avec la participation des acteurs (enseignants, étudiants...), en vue de prises de décisions concernant la conception, l'organisation, le fonctionnement et le financement des dispositifs. Les publications de langue anglaise différencient les termes désignant les deux champs : pour le premier champ, on utilise en général le terme *assessment* (e.g., *assessment for learning, personnel assessment*) et pour le second l'expression *program evaluation*¹.

1. Le terme *assessment* est parfois utilisé au sein d'une évaluation de dispositifs ; on pourrait dire, par exemple, *PISA evaluates educational systems through the assessment of students' knowledge and skills*.

Dans la suite du chapitre, j'aborderai à l'intérieur du premier champ seulement le cas de l'évaluation des apprentissages de personnes en formation, que je contrasterai avec l'évaluation des dispositifs de formation. Je tenterai de montrer en quoi, dans chacun de ces champs de l'évaluation, l'activité évaluative est orientée par des paradigmes de référence qui répondent à des exigences de la pratique très différentes, des exigences que je considère *in fine* comme étant inconciliables. Le caractère inconciliable des exigences de la pratique d'évaluation dans les deux champs ne signifie pas que les champs eux-mêmes soient inconciliables. Les deux champs doivent cohabiter, voire s'articuler dans le fonctionnement des systèmes éducatifs, même si chacun possède sa dynamique propre non réductible à celle de l'autre champ. J'y reviendrai dans la conclusion.

Les paradigmes considérés dans ce chapitre sont ceux qui portent sur *l'activité évaluative* (les finalités visées, les démarches entreprises, par la personne ou l'instance qui effectue l'évaluation), et non pas les cadres théoriques qui éclairent les phénomènes qui sont évalués, à savoir les théories d'apprentissage, d'enseignement et de didactique dans le champ de l'évaluation des apprentissages ; les théories de gestion, d'ingénierie de la formation, d'innovation et de gouvernance dans le champ de l'évaluation des dispositifs. Cette deuxième catégorie de théories est essentielle à la conduite des évaluations, mais son traitement dépasserait le cadre de ce chapitre.

REMARQUES PRÉLIMINAIRES

Au préalable, je me permets de relever deux difficultés que je rencontre en lisant certains écrits sur l'évaluation en éducation. La première concerne les textes où l'auteur tient un discours général sur l'évaluation sans différenciation des champs : parfois le discours ne concerne qu'un champ mais l'auteur en parle comme si son propos était valable pour toute activité évaluative ; d'autres fois, l'auteur juxtapose des idées ou des observations provenant des deux champs d'évaluation sans tenir compte de leurs spécificités, ce qui entraîne, à mon sens, une troublante confusion d'exposition. Je suis perplexe aussi face à l'emprunt de termes à la philosophie et à l'épistémologie sans véritable discussion de leur transposition dans le contexte de l'évaluation en éducation. Par exemple, quel rapport y a-t-il entre la conception de Foucault (1966) de l'épistémè (de la Renaissance, de l'âge classique, de la modernité...) et l'usage du terme par Bonniol et Vial (1997) pour parler de l'épistémè de l'évaluation comme mesure, l'épistémè de l'évaluation comme gestion, l'épistémè de l'évaluation comme problématique du sens ?

On peut interroger les fondements épistémologiques de l'évaluation selon de multiples perspectives (philosophiques, sociologiques, psychologiques) dotées d'appareils conceptuels importants. Plutôt qu'un survol de diverses perspectives, ce chapitre sera centré sur le concept de paradigme et

son intérêt pour la conceptualisation et la conduite de l'évaluation en éducation. Dans la deuxième édition du célèbre ouvrage de Kuhn (1970/2008), *La structure des révolutions scientifiques*, le lecteur croise, au fil des pages, plusieurs utilisations du terme paradigme. La description première du paradigme semble désigner une sorte de *Weltanschauung* scientifique qui, à l'image de la physique newtonienne, s'appuie sur des « découvertes scientifiques universellement reconnues » (p. 11) et comprend « des lois, des théories, des applications et des dispositifs expérimentaux » qui définissent la nature des problèmes à traiter et des solutions à rechercher dans un champ scientifique (p. 30). Dans la postface de la deuxième édition, Kuhn propose le terme « matrice disciplinaire » pour cette vision générale (p. 248) et définit le paradigme comme un exemple prototypique (*exemplar*) – la seconde loi de Newton est citée en exemple – qui constitue une référence partagée, unanimement admise par les membres d'une communauté scientifique, assurant ainsi l'unification de leurs travaux. Au-delà de ces variations de définition, la conception de Kuhn tend à placer les sciences sociales et humaines dans une phase de développement préparadigmatique en raison de leur éclatement en une multitude de perspectives souvent divergentes. Cela n'a toutefois pas empêché les spécialistes de ces disciplines, y compris en sciences de l'éducation, d'adopter très largement le concept de paradigme en revendiquant l'idée d'une science multiparadigmatique, caractérisée par la coexistence de paradigmes alternatifs (Masterman, 1970). Pour chacun des deux champs de l'évaluation traités dans ce chapitre, je tenterai d'identifier les principaux paradigmes qui éclairent l'évaluation en éducation. Mes propos seront centrés sur les tendances actuelles sans tracer un historique de l'évolution des paradigmes. J'emploierai souvent l'expression « paradigme de référence » car les paradigmes qui orientent l'activité évaluative ou qui nous aident à la comprendre ont été généralement élaborés en dehors du domaine de l'évaluation, même si les travaux de ce domaine ont apporté de nouveaux éclairages.

POINTS DE RENCONTRE ENTRE LES CHAMPS DE L'ÉVALUATION

Malgré les spécificités, il existe des points de rencontre entre l'évaluation des apprentissages en formation et l'évaluation des dispositifs de formation. Les opérations constitutives du processus évaluatif, qui se déroulent de manière récursive plutôt que linéaire, sont semblables dans les deux champs : définition des objets et des finalités de l'évaluation à réaliser, recueil d'informations relatives aux objets, interprétation des informations recueillies en leur conférant une signification, formulation de jugements conduisant à des prises de décisions, à des régulations de l'action. La référence à des valeurs (pédagogiques, scientifiques, politiques, éthiques) – qui guident le choix des objets et finalités, la manière de formuler les interprétations, jugements,

régulations, décisions – est toujours au cœur de l'évaluation, qu'il s'agisse de l'évaluation dans la formation (Hadji, 1989) ou de l'évaluation portant sur un dispositif de formation (Shadish, Cook & Leviton, 1991). De plus, bon nombre de concepts développés dans un champ ont été transposés à l'autre, tel l'exemple bien connu de la distinction sommative-formatrice proposée par Scriven (1967) dans l'évaluation des dispositifs et transposée par Bloom, Hastings et Madaus (1971) à l'évaluation des apprentissages en formation.

Je tenterai de démontrer cependant qu'au-delà de ces points de rencontre, les deux champs d'évaluation se différencient de manière fondamentale du point de vue des acteurs impliqués, des objets évalués, des finalités visées par l'évaluation, et appellent par conséquent une différenciation des paradigmes de référence aptes à orienter leurs démarches. J'aborderai en premier le champ de l'évaluation des dispositifs de formation car les conceptions et méthodes développées dans ce domaine sont souvent citées – à tort, je pense – comme modèles des démarches à instaurer dans l'évaluation des apprentissages en situation de formation.

ÉVALUATION DES DISPOSITIFS DE FORMATION

Le terme « dispositif de formation » est utilisé au sens large pour désigner une configuration de conditions (matérielles, organisationnelles, sociales, symboliques) dans lesquelles une activité de formation se déroule. Il recouvre des entités dont la nature et l'échelle varient : programmes, méthodes et outils didactiques, projets d'intervention, innovations, établissements, systèmes éducatifs (Figari, 2008). Bien qu'il existe des différences entre l'évaluation d'un(e) programme/méthode/projet bien circonscrit(e) et l'évaluation à l'échelle internationale des systèmes éducatifs (comme dans les enquêtes PISA), plusieurs caractéristiques communes de ces évaluations les distinguent nettement des évaluations effectuées par un formateur.

Caractéristiques de l'évaluation des dispositifs de formation

La caractérisation qui suit est une schématisation qui s'applique aux modalités d'évaluation relativement formalisées donnant lieu à des traces tangibles (rapports ou autres produits), même si une part du processus d'évaluation reste toujours implicite.

- *Acteurs* : L'évaluation des dispositifs est menée en général par des agents spécialisés (chercheurs et/ou administrateurs) qui adoptent des démarches, variables selon le contexte et le but de l'évaluation, permettant l'implication des publics concernés dans le processus d'évaluation. Ces publics comprennent les acteurs intervenant directement dans le

dispositif (formateurs, étudiants, directeurs d'établissement, etc.) et divers *stakeholders* (détenteurs d'enjeux) plus éloignés : par exemple, parents d'élèves, représentants d'instances politiques, groupements citoyens. Les agents chargés d'une évaluation sont parfois associés aux prises de décisions résultant de l'évaluation, mais ils ne sont pas en règle générale les principaux responsables de ces décisions.

- *Objets* : L'évaluation d'un dispositif de formation peut porter sur des objets multiples : effets de la formation, processus et démarches qui caractérisent la formation, significations attribuées aux activités de formation par les différents acteurs, notamment. L'appréhension des objets exige toujours des prises d'informations sur des unités élémentaires (par exemple, les compétences des individus qui participent à la formation, les représentations des formateurs des finalités de la formation), mais les interprétations et conclusions portent forcément sur des entités plus larges : par exemple, les compétences de groupes ou sous-groupes d'apprenants (tendances générales, variations) ; une catégorisation des représentations des formateurs. Même dans une évaluation par études de cas, on voudrait que celles-ci soient suffisamment prototypiques (et/ou atypiques) pour apporter un éclairage plus large sur la nature du dispositif évalué.
- *Finalités* : L'évaluation doit apporter des informations qui permettront des décisions d'ordre politique concernant la conception, l'organisation, le fonctionnement, le financement du dispositif dans son ensemble ou de certaines de ses composantes majeures. Les décisions ne concerneront ni les individus (étudiants, formateurs), ni les unités élémentaires du dispositif (par exemple, activités didactiques réalisées par tel ou tel formateur). Prenons le cas de l'évaluation d'un dispositif d'appui aux établissements scolaires dans des zones économiquement défavorisées. L'évaluation devra éclairer la politique à adopter pour la distribution et l'utilisation des ressources, les critères pour désigner les établissements concernés, les écueils et effets pervers qu'il faudrait éviter, les moyens pour dynamiser la participation des acteurs, mais l'évaluation ne servira pas à la prise de décisions concernant les parcours individuels des élèves ou les démarches quotidiennes de gestion d'un établissement.

Paradigmes de référence pour l'évaluation des dispositifs de formation

Il existe une littérature volumineuse en anglais sur les paradigmes susceptibles d'orienter la recherche en éducation, dont fait partie la « recherche évaluative » portant sur les dispositifs de formation. Deux références majeures sont le volume édité par Guba (1990) et l'ouvrage de Mertens (2010). Selon la synthèse de Mertens, un paradigme de recherche ou d'évaluation est défini par quatre ensembles de croyances : 1) des croyances ontologiques

concernant la nature de la réalité ; 2) des croyances épistémologiques concernant la nature de la connaissance et ses origines ; 3) des croyances méthodologiques concernant les démarches à privilégier dans l'élaboration des connaissances scientifiques ; 4) des croyances axiologiques concernant la dimension éthique de la recherche ou l'évaluation. Sur cette base et en se référant à un grand nombre d'autres écrits sur la question, Mertens différencie quatre paradigmes qui orientent les travaux actuels de recherche et d'évaluation en éducation, à savoir les paradigmes « post-positiviste, constructiviste, transformatif et pragmatique » (p. 11). Pour cette auteure, comme pour Guba et d'autres, le concept de paradigme est utilisé dans le premier sens formulé par Kuhn (1970/2008) ; il désigne une vision générale de la manière de concevoir et de conduire la recherche, y compris la recherche évaluative. Il englobe des croyances épistémologiques mais aussi des préoccupations d'ordre ontologique, méthodologique, axiologique. Le paradigme est donc positionné, dans cette conception, en cadre de pensée plus large que l'épistémologie.

Un examen approfondi de chaque paradigme décrit par Mertens (2010) n'est pas possible dans ce chapitre, mais je relèverai – de manière succincte et forcément simplificatrice – certains traits majeurs. Le paradigme post-positiviste se distancie du positivisme par ses concepts de réalisme émergent et d'objectivation par approximations successives, mais retient la visée générale d'une modélisation scientifique des phénomènes étudiés. Le paradigme constructiviste est fondé sur la notion (relativiste) de réalités multiples, socialement construites, à découvrir à travers les interactions entre chercheurs et participants et la confrontation de leurs compréhensions des phénomènes en jeu. Le paradigme transformatif s'appuie sur un réalisme critique qui met un fort accent sur la construction historique et contextuelle des relations de pouvoir entre les acteurs concernés par la recherche. Le paradigme pragmatique est axé sur la résolution de problèmes par des approches variées, adaptées à la réalité (perçue ou vécue) de chaque contexte étudié. Firestone (1990) présente une excellente discussion critique des paradigmes susceptibles de guider la recherche évaluative, leurs apports et surtout leurs limites face aux exigences de la pratique sur le terrain. Pour cet auteur, l'évaluation d'un dispositif implique souvent un jonglage entre plusieurs exigences – scientifiques, culturelles, politiques – inhérentes au contexte dans lequel le dispositif s'inscrit.

Parmi les classifications dans la littérature francophone sur l'évaluation, celle de Rodrigues (2006) recoupe partiellement les quatre paradigmes précités. L'évaluation « objectiviste » relèverait du paradigme post-positiviste (voire du positivisme classique selon la description de Rodrigues) et une évaluation « subjectiviste » du paradigme constructiviste. La perspective d'une évaluation « dialectique et critique » rejoindrait le paradigme transformatif. Des questions pragmatiques liées à l'utilisation de l'évaluation traverseraient les trois paradigmes cités par Rodrigues.

Certains auteurs proposent une centration plus forte sur les spécificités de l'évaluation des dispositifs en évitant son assimilation à une forme de recherche scientifique. C'est le cas d'Owen et Rogers (1999) qui, dans une perspective associant pragmatisme et réalisme émergent, identifient cinq « mini-paradigmes » d'évaluation ayant des problématiques et méthodes spécifiques. Il s'agit d'évaluations définies comme : proactive, clarificative, interactive, de monitoring, d'appréciation d'impact. L'ouvrage de Shadish *et al.* (1991) propose de structurer une « théorie » de l'évaluation des dispositifs (*programs* dans leur terminologie) comprenant quatre composantes : une composante liée aux savoirs à produire (avec des dimensions ontologique, épistémologique, méthodologique), mais aussi des composantes liées aux valeurs véhiculées par le dispositif à évaluer, aux utilisations des apports de l'évaluation et aux pratiques sociales et professionnelles de l'évaluation. Tout en évitant le terme paradigme, les auteurs de cet ouvrage effectuent une analyse qui s'approche de la deuxième signification de Kuhn (1970/2008) du paradigme comme *exemplar*. Ils analysent, à la lumière des composantes de leur cadre théorique, les approches d'évaluation développées par plusieurs grands concepteurs du champ aux États-Unis (dont Campbell, Cronbach, Scriven, Stake, Weiss). L'analyse de chaque approche apporte une réflexion approfondie sur la possibilité de considérer l'évaluation des dispositifs comme un champ autonome, distinct de la recherche scientifique. Pour Shadish *et al.*, l'évaluation d'un dispositif est toujours imbriquée dans un contexte et dans un processus de prise de décisions d'ordre politique (au sens *policy*), sans que le rôle de l'évaluateur (personne ou instance) se confonde avec celui du décideur politique. L'évaluation exige une expertise spécifique : la maîtrise de divers outils scientifiques et la capacité de les adapter aux finalités évaluatives ; des compétences de communication et de négociation avec les acteurs en présence. Elle constitue dès lors une pratique spécialisée, comme le relève le sous-titre du livre : *Theories of practice*.

Dans la littérature francophone sur l'évaluation, Barbier (2001) a aussi questionné le « paradigme dominant » qui tend à « assimiler travail d'évaluation et travail scientifique » (p. 358). En évoquant, sans le développer, un « autre paradigme » qui serait de considérer « l'acte d'évaluation comme un acte social au même titre que les actes qu'il se donne pour objet » (p. 359), il va plus loin que Shadish *et al.* (1991). Ces derniers considèrent l'acte d'évaluer comme une expertise spécialisée imbriquée certes dans des réseaux sociaux, mais néanmoins différenciée des actes (de formation, de gestion, etc.) qui sont l'objet de l'évaluation.

ÉVALUATION DES APPRENTISSAGES EN SITUATION DE FORMATION

Ce champ concerne l'évaluation des apprentissages des personnes en formation (élèves, étudiants, stagiaires...) par le formateur (l'enseignant ou une équipe d'enseignants) chargé de conduire les activités de formation.

Les conceptions actuelles de la formation préconisent l'implication active des apprenants dans le processus d'évaluation, à travers des démarches d'autoévaluation, d'évaluation mutuelle entre pairs, de coévaluation avec le formateur (Allal, 1999). Aussi, dans la suite du chapitre, l'évaluation « en formation » inclura-t-elle à la fois les évaluations effectuées directement par le formateur et celles instaurées par le formateur dans lesquelles les personnes en formation jouent un rôle actif.

Caractéristiques de l'évaluation des apprentissages en formation

La caractérisation qui suit soulignera surtout les points de différence avec l'évaluation des dispositifs de formation.

- *Acteurs* : Dans l'évaluation des dispositifs (quel que soit le paradigme de référence), la personne (ou l'équipe) chargée de l'évaluation rencontre périodiquement les acteurs de la formation, mais leurs échanges se situent dans le cadre de l'évaluation à mener et ne relèvent pas (en règle générale) d'une autre mission qu'ils sont censés accomplir ensemble. En revanche, un formateur qui évalue les apprentissages de personnes en formation, avec une participation plus ou moins active de leur part, côtoie régulièrement ces personnes dans le cadre de son rôle premier qui est de conduire les activités de formation. L'évaluation est une dimension importante mais néanmoins subsidiaire des relations qui les lient. Il en va de même pour les relations que le formateur entretient avec d'autres personnes qui sont (subsidièrement) coresponsables ou destinataires des évaluations : collègues, inspecteurs, parents d'élèves. Par ailleurs, le formateur doit généralement conjuguer les rôles d'évaluateur et de décideur, ce qui est rarement le cas dans l'évaluation des dispositifs.
- *Objets* : Selon les visées de la formation, l'évaluation des « apprentissages » peut porter sur plusieurs types de transformations (cognitives, affectives, sociales, psychomotrices) constatées pendant et au terme d'une étape de formation. Le formateur a la responsabilité de mettre en œuvre des évaluations portant sur les apprentissages réalisés par chaque individu qui participe aux activités de formation. Bien que les règles de l'institution amènent le formateur parfois à évaluer les acquis d'un groupe (par exemple, calculer la moyenne du groupe-classe à une épreuve externe afin de l'indiquer dans le livret ou d'en discuter dans un conseil de classe), l'évaluation vise principalement à appréhender les parcours d'apprentissage *singuliers* des personnes en formation.
- *Finalités* : Le formateur effectue des évaluations qui répondent à une demande institutionnelle et sociale (évaluations sommatives/certificatives, pronostiques), ainsi que des évaluations servant à adapter et à réguler les processus d'enseignement/apprentissage (évaluations formatives).

Ces évaluations ont une caractéristique fondamentale : elles servent à des décisions concernant les individus qui participent à la formation en cours. Concrètement, le formateur doit décider, pour chaque personne en formation, quelles régulations envisager, quelles appréciations attribuer au terme d'une étape, quelles recommandations fournir pour l'avenir. Si les informations provenant de l'évaluation aident parfois le formateur à ajuster la formation qui sera offerte aux volées futures, leur finalité première est de fonder les décisions les plus justes possible à l'égard des personnes actuellement en formation.

Paradigmes de référence pour l'évaluation des apprentissages en formation

Les travaux de docimologie, de pédagogie expérimentale et de psychométrie appliquée ont longtemps considéré que les démarches d'évaluation du formateur devaient reproduire – à une échelle plus réduite et sur la base d'instruments moins perfectionnés – les démarches valorisées dans la recherche évaluative. On croyait notamment que l'amélioration des évaluations effectuées par les enseignants proviendrait du rapprochement de leurs pratiques des principes scientifiques des théories de la mesure. Les règlements adoptés dans certains systèmes scolaires continuent à faire perdurer cette perspective, qui reflète non seulement une incompréhension du rôle de l'évaluation dans les processus de formation, mais aussi une méconnaissance des fondements scientifiques de la mesure.

Cela fait plus de 60 ans que Scates (1943) a décrit l'activité évaluative des enseignants comme « *interplay between objectivity and judgement* » (p. 6) ayant une autre logique que les démarches scientifiques de la mesure. Des recherches sur les pratiques d'évaluation en formation montrent en effet que le formateur navigue entre des démarches d'objectivation et la formulation de jugements interprétatifs concernant la progression des apprentissages de chaque élève (Allal, 1988 ; Allal & Mottier Lopez, 2008 ; Wyatt-Smith, Klenowski & Gunn, 2010). En cela, les évaluations réalisées par le formateur sont semblables à celles effectuées par d'autres professionnels (médecins, infirmières, assistants sociaux, conseillers en emploi, etc.) qui ont pour tâche d'agir face à la situation singulière d'un individu en tenant compte d'informations diverses, les unes standardisées, les autres hautement personnalisées. Le formateur doit gérer des dimensions collectives du processus d'évaluation : créer et appliquer des procédures d'évaluation à l'ensemble des personnes en formation ; tenir compte de règles convenues avec ses collègues ou imposées par le système de formation. Mais au bout de ce processus, il doit entreprendre l'action la plus juste possible par rapport au devenir de chaque apprenant. Cette exigence déontologique peut expliquer, à mon avis, pourquoi les conceptions du formateur de l'équité

en matière d'évaluation et des risques liés aux erreurs d'appréciation différente de celles typiquement retenues dans les évaluations portant sur les dispositifs (Allal, 2009).

Les paradigmes de référence de ce champ d'évaluation n'ont pas fait l'objet de synthèses systématiques comme celles citées pour l'évaluation des dispositifs. On peut identifier toutefois deux approches. La première est de considérer que la réflexion et l'action du formateur sont guidées par des croyances épistémologiques et ontologiques analogues à celles qui sous-tendent les paradigmes de la recherche évaluative. Si, à un certain niveau de généralité, on peut postuler que toute activité humaine reflète des croyances implicites d'ordre épistémologique et ontologique, l'exploitation faite de ces concepts pour rendre compte de l'activité des enseignants est souvent peu convaincante. Par exemple, selon Schraw et Olafson (2008), un enseignant « réaliste » sur le plan ontologique croit que la réalité est identique pour tout le monde et, par conséquent, ne tiendra pas compte des différences entre enfants, tandis qu'un enseignant « relativiste » qui croit en des réalités variables adaptera son enseignement aux besoins individuels des enfants. Ces affirmations reflètent un manque de compréhension de l'ontologie réaliste, notamment dans sa conception contemporaine du réalisme émergent (voir Greene & Caracelli, 1997). Un enseignant qui croit en une ontologie réaliste peut parfaitement penser que les différences entre enfants sont réelles, dans leurs manifestations et leurs effets, qu'elles résultent de facteurs multiples (en partie socialement construits), et qu'il faudrait adapter par conséquent son enseignement (ou évaluation) en fonction de ces différences. Sans développer ici la critique de cette approche, je ne pense pas qu'elle puisse en l'état éclairer l'activité évaluative du formateur.

Une autre approche est offerte par les recherches qui ont pour but d'identifier les logiques propres aux pratiques professionnelles en situation, et plus particulièrement celles ayant une visée clinique ou éducative. Ces recherches ont produit une grande diversité de modélisations qui ne forment pas un paradigme unifié mais qui contribuent néanmoins à une sorte de « perspective paradigmatique » axée sur une conception située des relations réflexives entre cognition individuelle et pratique sociale. Les travaux de Schön (1983/1994) sur la réflexion des professionnels dans et sur l'action montrent que les processus de raisonnement du praticien (faire des liens entre une situation inédite et des situations déjà vécues, imaginer des pistes d'action nouvelles) s'élaborent en regard d'un répertoire de pratiques reconnues dans la profession. D'après les études de l'activité clinique notamment dans les professions de la santé, la formulation d'un jugement évaluatif implique la mise en relation du singulier (tout ce que l'évaluateur sait de l'individu concerné) et du général (savoirs disciplinaires formels, normes et règles institutionnelles) afin de décider d'une action à la fois bien adaptée et bien contextualisée ; c'est notamment par la construction de « narratifs » que le professionnel donne sens aux liens

entre les informations à disposition et les voies d'action envisagées, avec leurs risques et leurs bénéfices potentiels (White & Stancombe, 2003). Un troisième apport vient des recherches sur le processus de *sensemaking* dans le champ de la psychologie sociale des organisations. Selon la synthèse établie par Weick, Sutcliffe et Obstfeld (2005), *sensemaking* est un processus continu par lequel le professionnel capte des éléments clés (*cues*) d'une situation, leur confère une signification rétrospective, les intègre dans un narratif plausible qui permet de comprendre la situation et d'organiser l'action à entreprendre ; c'est en même temps un processus social reflétant l'identité du professionnel, ses dialogues et transactions avec autrui, les contextes organisationnels et culturels de l'institution. Selon la conception située de Wenger (1998), l'identité du professionnel s'élabore et la signification de ses actions se négocie dans les multiples « communautés de pratiques » auxquelles il participe (par exemple, pour l'enseignant les communautés constituées par sa classe, son école, son association professionnelle, etc.). Un défi majeur pour le professionnel est alors la « réconciliation » des différentes significations conférées à ses actions dans différents contextes. Une pratique d'évaluation donnée (par exemple, l'attribution d'une note à un travail de groupe) peut être vue comme équitable dans la communauté-classe où elle a été élaborée mais comme inéquitable dans la communauté de l'école si celle-ci valorise fortement l'effort individuel. Malgré la diversité des orientations que je viens d'esquisser, elles soulignent toutes la réflexivité entre processus cognitifs et pratiques sociales et contribuent à ce titre à une perspective paradigmatique apte à analyser – ou à guider – l'activité d'évaluation du formateur en tant que professionnel opérant au sein d'une institution éducative.

EN CONCLUSION

Ce chapitre avait pour but de montrer que l'évaluation des dispositifs de formation, avec ses visées à la fois scientifique et politique, diffère de manière fondamentale de l'évaluation en situation de formation, avec sa dynamique clinique et sa finalité de prise en charge éducative. Les exigences de l'activité évaluative dans les deux champs d'évaluation sont liées aux rôles des acteurs qui conduisent l'évaluation, au niveau de généralité/singularité des objets évalués et aux finalités visées par l'évaluation. Ces exigences me paraissent *in fine* inconciliables car, prises ensemble, elles créent deux configurations de l'activité évaluative qui ne sont pas réductibles l'une à l'autre. Des exigences inconciliables appellent des paradigmes de référence qui ne sont pas seulement différents, mais de nature différente. Dans le cas de l'évaluation des dispositifs, les paradigmes doivent éclairer une activité évaluative qui réponde à des critères scientifiques et aux interrogations politiques des instances (*stakeholders*, décideurs) responsables des dispositifs. Dans le cas

de l'évaluation en formation, les paradigmes doivent éclairer une activité évaluative de nature clinique menée par un formateur dont la responsabilité première est l'encadrement éducatif des individus en présence. Je ne vois aucun intérêt à fusionner ces deux types d'activité dans une vision unifiée de l'évaluation qui estomperait leurs spécificités. L'évaluation peut très bien affirmer son pluralisme, comme le font bien d'autres domaines en sciences sociales et humaines.

Le terme « inconciliable » peut être vu comme excessif. Ne suffirait-il pas de dire que les exigences de l'activité évaluative dans les deux champs sont simplement très différentes ? Sans doute, d'autant plus qu'il existe des points de rencontre entre les deux champs de l'évaluation, comme cela a été indiqué dans la deuxième partie de ce chapitre. L'idée d'inconciliabilité est néanmoins utile, je pense, pour prévenir la tentation réductionniste présente depuis longtemps dans les écrits sur l'évaluation. Dans la tradition docimologique, mais aussi dans des travaux récents sur l'évaluation des systèmes éducatifs, l'évaluation effectuée par le formateur est souvent jugée à la lumière des critères scientifiques de l'évaluation des dispositifs et considérée en conséquence comme déficitaire et peu digne de confiance. Prenons un exemple, développé davantage ailleurs (Allal, 2009). On utilise couramment le terme « biais » pour décrire l'écart entre la note donnée par un formateur et celle (présumée juste) provenant d'un test fabriqué par un évaluateur/chercheur. Or, si l'on adopte les critères de l'activité évaluative du formateur, on pourrait parler de « biais » qui entachent le test standardisé car il ne tient pas compte de certains facteurs qui contribuent fortement à la réussite scolaire. Le même type d'argumentation s'appliquerait à l'autre champ d'évaluation : il serait tout aussi inadéquat d'aborder l'activité d'évaluation des dispositifs selon les exigences du travail du formateur. La personne ou l'équipe qui conduit l'évaluation d'un dispositif doit posséder une expertise reconnue sur le plan scientifique (quel que soit le paradigme de référence adopté) et ne peut être assimilée, à mon avis, à une sorte de formateur/conseiller qui encadre de manière clinique l'instance responsable du dispositif. Parler d'inconciliabilité nous met en garde non seulement contre des transpositions hâtives mais également contre toute tentative d'aménager une sorte de terrain neutre (ni scientifique, ni clinique) dans lequel les activités des deux champs se rejoindraient, mais perdraient les logiques propres de leurs pratiques.

Une forte différenciation des deux champs d'évaluation n'empêche toutefois pas leur cohabitation dans des systèmes éducatifs. Je pense qu'il est même nécessaire de promouvoir activement des mises en relations et des synergies entre les deux champs, afin d'assurer une action éducative de qualité. Mais la recherche des synergies sera productive à condition de reconnaître pleinement les spécificités (inconciliables) des deux champs.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. K. (1988). Quantitative and qualitative components of teachers' evaluation strategies. *Teaching and Teacher Education*, 4, 41-51.
- Allal, L. (1999). Impliquer l'élève dans le processus d'évaluation : Promesses et pièges de l'autoévaluation. In C. Depover & B. Noël (Éd.), *L'évaluation des compétences et des processus cognitifs* (pp. 35-56). Bruxelles : De Boeck.
- Allal, L. (2009). Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire. In L. Mottier Lopez & M. Crahay (Éd.), *Évaluations en tension : Entre la régulation des apprentissages et le pilotage des systèmes* (pp. 29-45). Bruxelles : De Boeck.
- Allal, L. & Mottier Lopez, L. (2008). Mieux comprendre le jugement professionnel en évaluation : Apports et implications de l'étude genevoise. In L. Lafortune & L. Allal (Éd.), *Jugement professionnel en évaluation : Pratiques enseignantes au Québec et à Genève* (pp. 223-239). Québec : Presses de l'Université du Québec.
- Barbier, J.-M. (2001). Quelques questions pour la recherche dans le domaine de l'évaluation. In G. Figari & M. Achouche (Éd.), *L'activité évaluative réinterrogée : regards scolaires et socioprofessionnels* (pp. 351-360). Bruxelles : De Boeck.
- Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York : McGraw Hill.
- Bonniol, J.-J. & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles : De Boeck.
- Figari, G. (2008). L'évaluation des dispositifs éducatifs. *Mesure et évaluation en éducation*, 31, 77-93.
- Firestone, W. A. (1990). Accommodation : Toward a paradigm-praxis dialectic. In E. G. Guba (Éd.), *The paradigm dialog* (pp. 205-224). Newbury Park, CA : Sage.
- Foucault, M. (1966). *Les mots et les choses : Une archéologie des sciences humaines*. Paris : Gallimard.
- Greene, J. C. & Caracelli, V. J. (Éd.). (1997). *Advances in mixed-method evaluation : The challenges and benefits of integrating diverse paradigms*. San Francisco : Jossey-Bass.
- Guba, E. G. (Éd.). (1990). *The paradigm dialog*. Newbury Park, CA : Sage.
- Hadji, C. (1989). *L'évaluation, règles du jeu : Des intentions aux outils*. Paris : ESF.
- Kuhn, T. S. (2008). *La structure des révolutions scientifiques* (L. Meyer, trad., 2^e éd.). Paris : Flammarion. (Original publié 1970).
- Masterman, M. (1970). The nature of a paradigm. In J. Lakatos & A. Musgrave (Éd.), *Criticism and growth of knowledge* (pp. 59-89). Cambridge : Cambridge University Press.
- Mertens, D. M. (2010). *Research and evaluation in education and psychology* (3^e éd.). Los Angeles : Sage.

- Owen, J. M. & Rogers, P. J. (1999). *Program evaluation : Forms and approaches*. London : Sage.
- Rodrigues, P. (2006). Logiques d'évaluation en relation avec des conceptions élargies du monde et de la connaissance. In G. Figari & L. Mottier Lopez (Éd.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie* (pp. 193-201). Paris : L'Harmattan.
- Scates, D. E. (1943). Differences between measurement criteria of pure scientists and of classroom teachers. *Journal of Educational Research*, 37, 1-13.
- Schön, D. (1994). *Le praticien réflexif : À la recherche du savoir caché dans l'agir professionnel* (J. Heynemann & D. Gagnon, trad.). Montréal : Éditions Logiques. (Original publié 1983).
- Schraw, G. J. & Olafson, L. J. (2008). Assessing teachers' epistemological and ontological worldviews. In M. S. Khine (Éd.), *Knowing, knowledge and beliefs : Epistemological studies across diverse cultures* (pp. 25-44). New York : Springer.
- Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series on Evaluation*, 1, 39-83.
- Shadish, W. R., Cook, T. C. & Leviton, L. C. (1991). *Foundations of program evaluation : Theories of practice*. Newbury Park, CA : Sage.
- Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16, 409-421.
- Wenger, E. (1998). *Communities of practice : Learning, meaning, and identity*. Cambridge : Cambridge University Press.
- White, S. & Stancombe, J. (2003). *Judgement in the health and welfare professions*. Maidenhead, UK : Open University Press.
- Wyatt-Smith, C., Klenowski, V. & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment : A study of standards in moderation. *Assessment in Education : Principles, Policy & Practice*, 17, 59-75.

À la recherche de paradigmes unificateurs ou de paradigmes intégrateurs ?

Jean-Marie De Ketele
Université catholique de Louvain

En essayant de mettre en cohérence les composantes de l'épistémologie, jugées essentielles par des auteurs comme Lalande (1968), Piaget (1970), Bachelard (1966), Kuhn (1972), Popper (1982), Besnier (2005), l'École supérieure de l'éducation nationale, de l'enseignement supérieur et de la recherche (2009) et Rebber (2010), nous définissons l'épistémologie de l'évaluation comme l'étude critique des processus de constitution des connaissances sur l'évaluation, des usages qui en ont été faits et de la valeur qui leur a été attribuée dans des situations sociales et historiques spécifiques. La prise en compte du caractère situé permet de comprendre que les études épistémologiques sur les développements de l'évaluation ont dégagé, à travers le temps, des paradigmes de l'évaluation qui ont été pris comme des modèles privilégiés à suivre par certaines catégories d'acteurs à des époques ou dans des lieux donnés. En effet, un paradigme est un ensemble de croyances et de présupposés tant théoriques que méthodologiques, édifié à une époque donnée et qui va servir de credo (Kuhn, 1972) et de modèle (Le Moigne, 1999) pendant un certain temps à des chercheurs et des utilisateurs. Les modèles d'évaluation attachés à un paradigme particulier sont des représentations conceptuelles formelles permettant de rendre intelligibles (aspect théorique), opérationnalisables (aspect méthodologique) et éthiques (aspect dégagé comme important par Rebber [2010]) les processus d'évaluation à conduire. Qui dit paradigmes différents dit modèles différents, dit conceptualisations et usages différents. Cette diversité donne l'impression d'un « éclatement de l'évaluation ». Nous pensons qu'il s'agit bien plus d'un élargissement progressif de l'évaluation, comme l'évoquait

déjà Cardinet en 1977. Nous émettons l'hypothèse que l'étude critique de cet élargissement nous permettrait peut-être de dégager à travers le temps un processus d'intégration des connaissances provisoires développées à travers les différents paradigmes identifiés. Cette étude des processus d'intégration nous amènerait à dégager une configuration de connaissances théoriques, méthodologiques et éthiques rendant intelligibles les différents paradigmes et modèles de l'évaluation, les considérant non pas simplement comme des moments historiques mais comme des strates émergeant progressivement dans un contexte donné, prenant de l'importance au fil du temps pour voir ensuite leur impact diminuer au profit d'une autre strate, sans jamais disparaître pour autant¹. Il s'agirait bien là d'une épistémè de l'évaluation. Telle est la thèse que nous tenterons de valider ou d'invalider en prenant comme corpus les écrits des experts du domaine de l'évaluation.

Comme l'épistémologie est l'étude des processus de constitution des connaissances, nous procéderons en plusieurs temps. En premier lieu, nous essaierons d'analyser dans leur contexte historique le développement de différents modèles d'évaluation, ce qui permettra de resituer ce que nous avons appelé les différents « paradigmes de l'évaluation » (De Ketele, 1993) que nous réinterpréterons à la lumière d'une classification en quatre modèles ou paradigmes princeps et d'un processus de construction de strates qui s'emboîtent et se superposent, et non pas comme des périodes successives. Nous aurons aussi l'occasion de dégager, pour chacun des paradigmes, les fondements conceptuels et idéologiques sous-jacents. Ensuite, nous tenterons de voir si un paradigme unificateur, à la fois d'élargissement (voir les travaux de Cardinet) et d'intégration, permet de mieux comprendre le développement des travaux sur l'évaluation.

QUATRE PARADIGMES PRINCEPS ET QUATRE ÉPISTÉMOLOGIES

Nous identifions quatre paradigmes princeps de l'évaluation en croisant deux axes d'analyse : d'une part, les savoirs produits ; d'autre part, les normes de référence. Nous pouvons distinguer essentiellement deux types de savoirs produits par les démarches évaluatives : les savoirs en termes de jugement (voir les formalisations de Noizet & Caverni [1978] et de Bressoux & Pansu [2003]) et les savoirs d'action (voir les formalisations de Stufflebeam *et al.* [1980]). Cette première distinction est axée sur le produit du processus évaluatif : i) le produit recherché est un jugement (souvent formulé sous forme de note ou d'une mesure plus ou moins sophistiquée) établissant un bilan susceptible d'être communiqué à une audience plus ou moins étroite (les

1. Nous avons déjà utilisé cette métaphore des strates biologiques pour mettre en évidence comment le statut du savoir a évolué au cours des siècles (De Ketele, 2000).

élèves ou les parents, par exemple) ou plus large (les classements internationaux, par exemple) ; ii) le produit recherché est un savoir d'action quand le processus évaluatif a pour finalité directe de prendre des décisions pour l'action. Si le jugement évaluatif peut conduire à prendre des décisions pour l'action, le processus qui y a conduit n'a pas été conçu prioritairement pour cela ; par ailleurs, si le processus évaluatif qui a pour finalité première de produire un savoir servant l'action comprend des démarches de jugement (choix des objectifs à privilégier, des critères à retenir, des informations à recueillir...), le produit final n'est pas un jugement à communiquer mais la détermination d'une décision à prendre pour l'action. Que le produit soit un jugement ou un savoir d'action, nous trouvons essentiellement deux types de normes de référence : les normes ou connaissances locales ; les normes ou connaissances globales que l'on appelle souvent des « standards ». Cette dernière distinction rejoint celle de Figari (2006) qui parle de connaissances locales *versus* générales, si ce n'est qu'il les considère comme des résultats produits alors que nous les envisageons ici comme des connaissances de référence. Non seulement le processus évaluatif produit des connaissances nouvelles sur l'objet, mais il ne part pas de rien ; il se fait à partir d'un référentiel, c'est-à-dire de connaissances qui peuvent être locales (ce que font beaucoup d'enseignants lorsqu'ils évaluent) ou de connaissances prescrites (connaissances globales : référentiel ou standards). Le croisement de ces deux axes (type de produits recherchés et type de normes de référence) nous fournit quatre paradigmes princeps.

Le paradigme de *l'intuition pragmatique* a pris naissance avec l'avènement formel de l'école (Beillerot, 2001) et s'est renforcé avec le phénomène de bureaucratization de celle-ci (Bonami, 1996). L'intuition pragmatique caractérise les pratiques les plus courantes de l'enseignant lorsqu'il évalue : il produit des jugements sur les performances des élèves sur la base de ses connaissances locales. Le paradigme de *l'évaluation mesure* prend sa source dans les recherches docimologiques (de Landsheere, 1980 ; Noizet & Caverni, 1978 ; Piéron, 1963) et s'est vu renforcé par les critiques émises suite aux recherches sociologiques (Bourdieu & Passeron, 1964 ; Crahay, 1996 ; Duru-Bellat, 2002 ; Perrenoud, 1984) sur la contribution de l'évaluation au maintien des inégalités sociales. Le paradigme de l'évaluation mesure, actuellement présent dans le mouvement de l'éducatrice largement inspiré par les psychométriciens, tend à produire des jugements objectivés en référence à des connaissances globales, si possible des standards internationaux établis par des experts. Le paradigme de *l'évaluation de la maîtrise des apprentissages* s'est construit sur la base des principales critiques adressées à l'évaluation mesure, selon lesquelles les jugements produits ne sont en général pas construits pour servir et réguler les actions du système éducatif, comme l'analyse des pratiques en témoigne. Le paradigme de l'évaluation de la maîtrise des apprentissages vise à produire des savoirs d'action en référence à des connaissances globales (si possible des standards internationaux) à travers la mise en œuvre des principes de

la pédagogie de maîtrise et des pratiques évaluatives centrées particulièrement sur la fonction formative (Allal, Cardinet & Perrenoud, 1979 ; Bloom, Hastings & Madaus, 1971 ; Scriven, 1967). Centré également sur l'apprentissage et la régulation, le paradigme de l'évaluation *pour* les apprentissages (*assessment for learning*, développé par l'*Assessment Reform Group* en 1999 et diffusé en langue française par Stobart en 2011) se distingue du précédent par « une interprétation culturelle spécifique de l'évaluation formative » (Stobart, 2011, p. 43) qui, selon cet auteur, s'inscrit dans la régulation interactive décrite par Allal et Mottier Lopez (2005). En s'appuyant sur les travaux francophones de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE), Allal et Laveault (2009) préfèrent appeler ce paradigme *évaluation-soutien d'apprentissage*, ce qui serait d'ailleurs à leurs yeux une traduction plus adéquate. En effet, les pratiques illustrant ce paradigme cherchent à générer des savoirs d'action en référence aux connaissances locales produites dans l'action et l'interaction entre l'apprenant et l'enseignant, en tenant compte de l'histoire contextualisée de l'apprentissage (Stobart, 2011) et avec l'intention de soutenir l'apprentissage et le parcours de la personne en formation (Allal & Laveault, 2009). Si la régulation interactive est privilégiée dans ce paradigme, les deux autres formes de régulation décrites par Allal et Mottier Lopez (2005) (régulations rétroactive et proactive) peuvent être associées au paradigme de l'évaluation de la maîtrise des apprentissages.

Afin de bien comprendre ces quatre paradigmes, il nous semble intéressant de produire une synthèse comparative des dimensions épistémologiques dégagées suite à une analyse des pratiques méthodologiques des évaluateurs appartenant aux quatre paradigmes princeps (voir tableau 1). Nous utilisons sept dimensions qui correspondent à des composantes clés de la méthodologie de l'évaluation : quel type de connaissance l'évaluateur veut-il produire ? En référence à quel type de normes ? Sur base de quelles informations ? Pour répondre à quelle question centrale ? Avec quel type de traitement de l'information et d'attribution de sens ? Pour quels destinataires privilégiés ? Avec quels concepts clés ?

Tableau 1 : Synthèse comparative des dimensions épistémologiques des quatre paradigmes

Dimensions épistémologiques	Évaluation intuitive pragmatique	Évaluation mesure	Évaluation de la maîtrise des apprentissages	Évaluation <i>pour</i> les apprentissages
<i>Connaissance produite</i>	Jugement sommatif local intuitif sur les élèves	Jugement sommatif sur une échelle de mesure objectivée	Savoirs d'action sur la maîtrise des apprentissages en référence à des standards globaux	Savoirs d'action locaux en référence aux progrès possibles des élèves

Dimensions épistémologiques	Évaluation intuitive pragmatique	Évaluation mesure	Évaluation de la maîtrise des apprentissages	Évaluation pour les apprentissages
<i>Normes ou savoirs de référence</i>	Les savoirs enseignés localement et les comportements des élèves	Des standards de référence, à condition qu'ils soient objectivables et mesurables sur une échelle unidimensionnelle	Les standards de référence prescrits traduits en objectifs de maîtrise	Les savoirs locaux susceptibles d'être appris par les élèves
<i>Dispositif de recueil d'informations</i>	Les interrogations, les tests, les examens, de même que les comportements des élèves vis-à-vis de l'enseignement et de l'étude	Des épreuves ponctuelles standardisées avec des items validés par des modèles psychométriques (TRI)	Des tests réguliers de maîtrise des apprentissages prescrits, y compris la maîtrise des prérequis	Des auto et des coévaluations des démarches et des produits de l'apprentissage mis en relation avec l'histoire des apprentissages
<i>Question centrale posée aux élèves</i>	Ont-ils étudié et appris comme je l'avais demandé ?	Sont-ils bien situés sur une échelle de performance ?	Ont-ils maîtrisé les standards de performances prescrits ?	Ont-ils pris conscience qu'ils ont progressé grâce aux démarches mises en œuvre et aux ressources mobilisées ?
<i>Traitement de l'information</i>	Sommatif intuitif	Sommatif objectif	Sommatif ou descriptif selon les cas	Interprétatif
<i>Destinataires</i>	Les acteurs qui ont un pouvoir sur le comportement de l'élève	Les acteurs qui ont un intérêt à comparer	Les acteurs internes et externes qui ont une mission de régulation	Les acteurs locaux, en particulier les élèves avec leurs enseignants
<i>Concepts clés</i>	Contrat didactique Contrat pédagogique	Mesure (docimologie et éduométrie) Quasi-marché	Pédagogie de maîtrise Évaluation formative proactive et rétroactive	Auto et coévaluation Régulation interactive et soutien

LES FONDEMENTS CONCEPTUELS ET IDÉOLOGIQUES DES QUATRE PARADIGMES

Pour appréhender de la façon la plus exhaustive possible les fondements épistémologiques conceptuels et idéologiques des quatre paradigmes principes, il nous semble important de les inférer à partir des caractéristiques méthodologiques et de leurs conséquences. Dans cette section, nous ne mettrons pas l'accent sur les concepts que les différents modèles ont en commun, comme la confrontation d'un référé à un référent ou la fonction de régulation de toute évaluation. Ce qui nous intéresse ici est ce qui les différencie. Ainsi, les référés et les référents varient d'un modèle à l'autre ; de même, la régulation a des visées différentes suivant les moments de l'évaluation et les personnes impliquées. C'est dans ce même esprit que nous envisagerons les autres dimensions épistémologiques de nature plus idéologiques.

L'évaluation intuitive pragmatique

Les études docimologiques (avec l'accent mis sur les biais) et sociologiques (avec la mise en évidence des inégalités sociales) sur les pratiques évaluatives des enseignants pouvaient induire un jugement négatif sur celles-ci. Avec sa contribution célèbre *La subjectivité blanchie ?*, Weiss (1984) a largement contribué à atténuer ce jugement en soulignant l'importance de l'intuition de l'enseignant et de la praxis. Les travaux ultérieurs sur le développement professionnel des enseignants lui donneront raison en dégageant le concept de *savoirs d'expérience* (Barbier, 1996). De même, de Landsheere (1980) relèvera toute une série d'études empiriques montrant que les évaluations intuitives des enseignants étaient souvent mieux corrélées à la réussite scolaire ultérieure que les résultats à des épreuves externes standardisées. Ceci nous amène à dire que le concept le plus caractéristique du paradigme de l'évaluation intuitive pragmatique est le concept de *contrat didactique* que Chevallard (1986) a contribué à populariser. Le pouvoir que l'évaluation donne à l'enseignant est utilisé par lui pour conformer les élèves au contrat didactique, éventuellement pour l'ajuster, en jouant sur les facteurs de motivation et d'estime de soi. Comme de nombreux travaux l'ont montré (voir par exemple Gilly, 1980), les enseignants ont intériorisé un modèle du « bon élève » (attentif lors de l'enseignement, participant activement, discipliné, respectueux, travailleur...) ; cette image s'accompagne souvent d'une conception innéiste selon laquelle il y aurait des élèves doués et d'autres qui le sont moins et que leur répartition serait en partie liée aux classes sociales. Les sociologues (comme Bourdieu et Passeron [1964] avec le concept de reproduction sociale ; ou Perrenoud [1984] avec le concept de fabrication de l'excellence scolaire) diront que les pratiques évaluatives des enseignants traduisent une idéologie du

déterminisme social. Cherchant à adapter l'élève à l'école et au contrat didactique, les enseignants adopteront donc davantage une évaluation avec une posture de contrôle et de sanction. On peut faire une analyse semblable lorsque l'on se situe au niveau méso de l'établissement scolaire ou de formation lors de l'examen des pratiques évaluatives en référence aux projets d'établissement et aux divers projets éducatifs et pédagogiques. On est encore ici dans une négociation contractuelle.

L'évaluation mesure

En réaction aux biais mis en évidence par les docimologues et aux injustices créées par l'évaluation, les spécialistes de la mesure tentent d'objectiver le jugement sommatif et s'appuient sur les théories et les méthodologies de la mesure et des tests, en prenant comme référence des standards externes dans leurs composantes objectivables. De Ketele et Gerard (2005) montreront que cette volonté d'objectivation à tout prix conduit à sélectionner les items qui sont les plus facilement évaluables, c'est-à-dire les connaissances à restituer fidèlement et les savoir-faire de base (que les enseignants appellent les « applications »). Or, les apprentissages les plus importants relèvent le plus souvent de tâches complexes contextualisées, comme la production de textes, la communication en situation, la résolution de problèmes, etc. Ce caractère réducteur montre bien que le concept d'échelle, à la fois de mesure et de performance, est à la base du paradigme avec la volonté de les rendre les plus isomorphes possible. Par ailleurs, qui dit échelle implique l'unidimensionnalité ou la recherche de plier la réalité évaluée à l'unidimensionnalité. Les évaluateurs vont donc se référer à des standards de performance susceptibles d'être objectivés et sommés (en général liés aux programmes prescrits pour les épreuves nationales) ; à des « compétences »² jugées utiles quels que soient les contextes dans les épreuves internationales, faute de pouvoir s'accorder sur un tronc commun. Les résultats seront des mesures qui conduisent à des jugements sommatifs comparatifs (entre élèves, entre classes, entre établissements, entre régions, entre pays, entre types de systèmes éducatifs, etc.). Le modèle produit et induit des palmarès dont les acteurs internes et externes des systèmes éducatifs s'emparent en les utilisant au mieux de leurs intérêts. Le système éducatif devient donc un *quasi-marché* que de nombreux travaux ont analysé (Dupriez & Vandenberghe, 2004 ; Vandenberghe, 1999). Et certains (comme Maroy, 2000, 2007) n'hésitent pas à parler d'idéologie néo-libérale et de recherche à tout prix d'*efficience*, c'est-à-dire du meilleur rapport entre résultats et ressources investies dans le

2. Dans l'article précité, De Ketele et Gerard (2005) montrent qu'en fait le modèle TRI n'évalue pas des compétences au sens plein du terme (c'est-à-dire des tâches complexes sur des situations complexes), mais une somme de tâches plus simples nécessaires à la maîtrise des compétences.

système éducatif. Par cet effort de « scientifier » (sous-entendu : le quantitatif est étroitement lié au caractère scientifique), on cherche à rendre l'école plus compétitive. Ici encore, on se trouve dans une posture de contrôle, mais légitimée par les compétences édumétriques des évaluateurs.

L'évaluation de la maîtrise des apprentissages

Les caractères à la fois réducteur (des objets essentiels ne sont pas pris en considération) et statique (une mesure ne donne pas d'indications concrètes pour les acteurs internes et externes) ont donc conduit au paradigme de l'évaluation de la maîtrise *des* apprentissages qui met en avant l'importance de produire des savoirs d'action utiles pour gérer et piloter les apprentissages, non seulement au niveau micro (les pratiques de classe) mais aussi méso (l'établissement scolaire) et macro (les autorités éducatives et politiques). Le pilotage en tant que processus de régulation à ces différents niveaux est donc le concept clé de ce modèle avec celui de maîtrise. Il s'agit de professionnaliser les enseignants, les responsables des établissements, les responsables nationaux. Comme le pilotage suppose des référents bien connus et si possible partagés, le modèle se réfère à des standards définis par des experts, non seulement au niveau national, mais de plus en plus aussi au niveau international comme le veut la globalisation. Ces standards vont servir de base à l'action et à l'évaluation certificative (la maîtrise des objectifs, base de la « pédagogie par objectifs » ou PPO) ; mais le pilotage aux différents niveaux nécessite en outre d'autres types d'évaluation pour orienter les apprentissages et améliorer les apprentissages en cours. On ne peut donc recourir exclusivement à des démarches sommatives ; il faut aussi faire appel à des démarches descriptives et rétroactives pour gérer la progression des apprentissages. Derrière ce modèle se révèle une idéologie de l'efficacité (maximiser le rapport entre les résultats observés de l'apprentissage et les résultats attendus) et de l'expertise (développer le professionnalisme des acteurs aux différents niveaux grâce aux savoirs d'action produits par les différents types d'évaluation). La finalité est donc bien d'adapter l'école (comme système) à l'apprentissage efficace. La posture de l'évaluateur est à la fois une posture de contrôle et de régulation (principalement rétroactive).

L'évaluation pour les apprentissages ou l'évaluation-soutien d'apprentissage

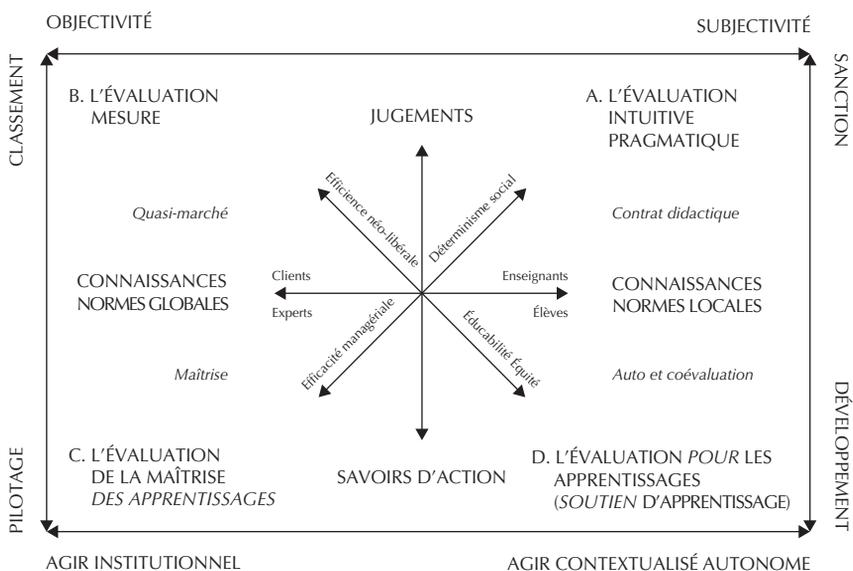
Si dans le paradigme précédent l'apprentissage est au cœur du processus, il n'en reste pas moins vrai que ce sont les acteurs en tant que professionnels qui sont les acteurs clés, malgré les nombreux discours qui insistent sur la place centrale de l'apprenant. Tout en ne remettant pas en cause l'importance du pilotage et de l'expertise, plusieurs lignées de travaux vont contribuer à

mettre l'accent sur les apprentissages eux-mêmes et sur l'évaluation *pour* les apprentissages – c'est-à-dire comme *soutien* d'apprentissage – et non plus simplement *des* apprentissages. Étudiant le foisonnement des réformes pédagogiques en tous genres, des chercheurs vont montrer l'inanité des efforts et des investissements consentis dans bien des cas. Ainsi, la méta-analyse de Lieberman (2005) montre que les changements mis en place ne produisent les effets escomptés que s'il y a une action à long terme sur les « pratiques de classe », qui ne se réduisent pas aux « pratiques des enseignants », mais concernent les apprentissages tels qu'ils se déroulent réellement, c'est-à-dire avant tout dans la tête et les comportements des élèves eux-mêmes, grâce à leurs interactions avec les ressources de l'environnement et à l'histoire de ces interactions. Les concepts de démarches et de ressources pour les apprentissages, ainsi que les concepts de progression et de progrès auto et corégulés deviennent centraux. Le pilotage n'est plus simplement un pilotage d'experts qui savent et qui conseillent, mais un pilotage d'experts accompagnateurs qui soutiennent, et où le premier expert est l'élève lui-même. Les travaux du colloque de l'ADMEE de Mons en 1998 (Depover & Noël, 1999) allaient déjà dans ce sens en mettant en évidence l'importance de la métacognition et des stratégies qui y conduisent. Il s'agit donc d'adapter l'environnement de l'apprentissage aux potentialités d'apprentissage et d'autorégulation de l'élève. Il ne s'agit donc plus d'une posture de contrôle pour réguler, mais d'une posture de reconnaissance dans l'esprit des travaux de Ricœur (2004) et que des auteurs comme Jorro (2006, 2009) et De Ketele (2010) ont précisée dans le cadre de l'évaluation : situer l'apprenant en tant que personne apprenante ; accepter l'apprenant dans ses démarches d'apprentissage ; valoriser l'apprenant en reconnaissant ses progrès ; être reconnaissant envers l'apprenant pour ce qu'il a fait en vue de progresser. On est donc ici dans une idéologie de l'éducabilité, bien loin de l'évaluation intuitive pragmatique, même si on accorde dans les deux cas une importance majeure au contexte local et à la motivation scolaire.

Suite à cette quadruple analyse, on peut tenter une synthèse comparative (voir figure 1). On y retrouve les deux axes précédemment cités (jugements *versus* savoirs d'action ; connaissances ou normes globales *versus* locales) et les quatre paradigmes développés. Notre analyse des fondements conceptuels et idéologiques nous amène à produire quatre nouveaux axes à la périphérie : i) les jugements sont caractérisés par et même valorisent « l'objectivité » *versus* la « subjectivité » ; ii) les savoirs d'action privilégient l'agir institutionnel *versus* l'agir contextualisé autonome ; iii) l'évaluation se met au service du classement *versus* le pilotage ; iv) l'évaluation est un outil qui sanctionne positivement ou négativement *versus* sert le développement de la personne de l'élève. À ces axes s'ajoutent deux axes idéologiques dessinés obliquement : v) un axe qui part de l'évaluation intuitive pragmatique qui sous-tend davantage une certaine forme de déterminisme social, et qui s'oppose et rejoint l'évaluation de la maîtrise des apprentissages qui relève

d'avantage de l'efficacité managériale aux différents niveaux du système éducatif ; vi) un axe qui part de l'évaluation mesure qui sous-tend davantage une idéologie de l'efficacité néo-libérale, et qui s'oppose et rejoint l'évaluation pour les apprentissages qui défend une idéologie de l'éducabilité et de l'équité. Sur cette même figure, on peut aussi placer les acteurs privilégiés en fonction des paradigmes, respectivement les enseignants, les clients, les experts et les élèves. Et parallèlement, les concepts prioritaires : le contrat didactique, le quasi-marché, la maîtrise des objectifs institutionnels, l'auto et la coévaluation dans les interactions.

Figure 1 : Synthèse comparative des fondements épistémologiques conceptuels, idéologiques et éthiques



VERS UN PARADIGME UNIFICATEUR ET INTÉGRATEUR ?

Pour tenter de répondre à cette question, il est intéressant de faire référence au monde de la recherche scientifique. Les écrits épistémologiques sur le processus de construction des connaissances scientifiques font allusion à des expressions comme « paradigme positiviste », « paradigme expérimental », « paradigme clinique »... et même « paradigme quantitatif » opposé à « paradigme qualitatif ». Si cela traduit bien la richesse de la recherche scientifique, De Ketele et Maroy (2011), de même que De Ketele et Paquay (2011), ont tenté de dégager un paradigme princeps unificateur derrière cette pluralité : toute recherche scientifique a pour fonction de produire

des connaissances validées par la communauté scientifique de l'époque et susceptibles d'être remises en cause par les communautés scientifiques ultérieures ; la connaissance scientifique produite est provisoirement validée lorsque le modèle d'intelligibilité proposé de la réalité étudiée est corroboré (non invalidé) par un ensemble de démarches et un raisonnement reconnu comme rigoureux, quelles que soient les approches (expérimentales, corrélationnelles, observationnelles) et démarches (quantitatives, qualitatives) adoptées. La pluralité est une richesse et une question de pertinence, et non d'opposition. Par ailleurs, la production des connaissances scientifiques est un processus intégrateur dans le sens où les nouvelles connaissances produites enrichissent et intègrent les connaissances antérieures non invalidées, éventuellement en les remodelant ou en précisant leurs limites de généralisation.

Il n'est évidemment pas question de confondre ici évaluation (les pratiques des évaluateurs) et recherche scientifique (dont notamment les recherches scientifiques sur l'évaluation)³, mais il est frappant de constater que toutes deux sont parcourues de paradigmes multiples. Peut-on espérer cependant proposer pour l'évaluation un paradigme princeps unificateur et intégrateur ?

En quelque sorte, l'évaluation en éducation a aussi pour but de produire des modèles d'intelligibilité, mais ceux-ci ne sont pas du même ordre, non seulement parce que leur légitimité n'est pas assurée par la communauté scientifique, mais parce que l'intelligibilité produite est censée servir certains acteurs du système éducatif et se produit en référence à des connaissances et normes différentes selon les acteurs concernés. Les jugements sommatifs intuitifs permettent de réguler le contrat didactique entre l'évaluateur et les élèves et se construisent en référence aux enseignements dispensés localement ; les jugements sommatifs objectivés permettent de réguler le système éducatif en tant que quasi-marché et se construisent en référence à des standards dans leurs composantes objectives et par des démarches permettant des comparaisons ; les savoirs d'action de l'évaluation de maîtrise *des* apprentissages permettent de piloter les apprentissages, et plus largement le système éducatif, en référence à des standards prescrits et grâce à des démarches expertes ; les savoirs d'action de l'évaluation *pour* les apprentissages et *soutien* d'apprentissage permettent de réguler les interactions de l'apprenant avec son environnement et s'élaborent en référence à des savoirs locaux ancrés dans une histoire et un contexte.

3. Dire que l'élève X maîtrise telle compétence, c'est une connaissance produite par un praticien de l'évaluation. Par contre, dire sur la base d'une analyse d'un corpus que la plupart des praticiens de l'évaluation se caractérisent par un paradigme de l'intuition pragmatique est une connaissance scientifique susceptible d'être remise en cause par la communauté scientifique des chercheurs en éducation.

Au-delà de la diversité des paradigmes et de leurs usages, on trouve aussi des éléments unificateurs derrière les démarches méthodologiques. Il s'agit chaque fois de la confrontation d'un référé à un référent, même si ceux-ci sont différents d'un paradigme à l'autre. On trouve donc toujours des démarches de recueil d'informations et des démarches de traitement des informations recueillies (certes différentes selon les cas). Les résultats du traitement de la confrontation entre le référé et le référent sont toujours le fruit d'un processus d'attribution de sens, mais différent selon les usages. Ce sont d'ailleurs les usages qui expliquent le caractère plus idéologique de l'évaluation, comparativement à la recherche, même si celle-ci n'en est pas exempte puisqu'elle est également une pratique sociale et située. Ce sont encore les usages qui permettent d'expliquer la multiplicité des postures de l'évaluateur.

Malgré ces quelques éléments unificateurs, il n'en reste pas moins vrai que les pratiques évaluatives et les paradigmes sous-jacents donnent l'impression d'un éclatement. À l'instar de Cardinet (1977), nous parlerions plus volontiers d'un processus historique d'élargissement : de l'intuition et de la subjectivité à la mesure objectivée pour réduire les biais des pratiques évaluatives ; de normes locales à des normes globales ; du classement selon une norme de population au pilotage selon des normes critériées de maîtrise des apprentissages ; de l'expertise des professionnels de l'éducation à l'expertise de l'apprenant.

Cet élargissement ne s'est pas effectué par périodes successives, mais par strates qui se superposent au fil du temps, continuent à coexister et tendent même, sur certains aspects, à intégrer des acquis antérieurs ou à dépasser certaines limites des strates précédentes. Le paradigme de l'évaluation intuitive pragmatique est né de la bureaucratisation du système éducatif et a pris rapidement de l'ampleur. La structuration de la recherche en éducation et le développement de la mesure dans de nombreuses disciplines ont fait émerger la seconde strate : l'évaluation est devenue mesure et objet de comparaison. Cette seconde strate a pris progressivement de l'importance sous l'influence des psychométriciens et des consortiums internationaux établissant des classements ; elle est loin d'être révolue. Mais rapidement, des protestations se sont élevées et, récemment, se sont intensifiées suite aux critiques contre les classements, considérés comme réducteurs et objet d'interprétations multiples ou de récupérations politiques. De cette configuration a émergé la troisième strate qui centre l'évaluation non plus simplement sur la base de jugements ponctuels sujets à interprétations diverses, mais sur des savoirs d'action qui permettent de piloter les apprentissages, et plus largement le système éducatif, celui-ci se fixant des objectifs de maîtrise à atteindre et comptant sur l'expertise professionnelle des acteurs pour réguler les actions dans ce sens. Une quatrième strate est venue se superposer aux

trois autres, fondée sur le constat que si l'apprenant est au centre des apprentissages dans la strate précédente, il reste un objet que l'expertise cherche à façonner en vue de la maîtrise des objectifs prescrits ; elle considère que l'apprenant est un sujet doté d'une expertise propre, et qu'aucun progrès n'est possible sans le recours à cette expertise qui se construit dans l'interaction avec l'environnement et dans une histoire. Dans les pays francophones, cette quatrième strate a pris progressivement de l'importance dans les travaux de l'ADMEE, mais reste curieusement encore peu répandue dans les pratiques de classe, contrairement aux pays anglo-saxons où elle gagne du terrain.

Les quatre strates coexistent donc et continueront à exister. L'existence d'un paradigme unificateur et intégrateur supposerait par conséquent d'admettre des raisonnements « si... alors » dans le cadre de deux principes. Le premier principe affirme que selon les acteurs et les usages, nous avons besoin de *différents modèles d'intelligibilité* qui ont chacun leur fonction et leurs limites : l'enseignant dans son contexte local a besoin de réguler le contrat didactique, sans lequel sa tâche n'est pas possible ; les instances nationales et internationales ont besoin de points de comparaison ayant pour fonction de susciter les prises de conscience, à condition cependant que les études éducatives parviennent à développer des modèles de validation pour évaluer des performances de niveau plus complexe ; les professionnels de l'éducation ont besoin de disposer de savoirs d'action pour réguler les apprentissages aux différents niveaux micro, méso et macro du système éducatif en relation avec la maîtrise des apprentissages jugés fondamentaux ; ces mêmes professionnels ont besoin de soutenir des évaluations situées au cœur même des apprentissages, où l'apprenant est l'évaluateur premier de ses propres apprentissages et prend conscience de la nécessité, à des moments clés, de recourir à l'autre (l'enseignant ou les pairs) pour rendre intelligibles les situations de blocage (Martinand [1986] parle « d'objectifs obstacles » ; mais il existe d'autres éléments à prendre en considération) et parvenir à progresser dans les apprentissages et la voie de l'autonomie (gage d'un apprentissage réussi). Le second principe repose sur le fait que ces « si... alors » sont tous indispensables, mais ont besoin d'un *pilotage systématique*, c'est-à-dire non seulement d'un *processus de référentialisation* (Figari, 1994), mais aussi et surtout d'un *processus d'attribution de sens par intégration* des différents modèles d'intelligibilité pour produire de nouveaux savoirs d'action. Il ne s'agit pas de confier cette tâche à une instance « supérieure » qui, forte de son expertise et de sa légitimité déclarée, répercuterait son analyse aux différents niveaux ; il s'agit d'un processus participatif fondé sur une posture de la reconnaissance, où chaque professionnel pourrait bénéficier de l'apport des différents modèles d'intelligibilité pour approfondir les modèles d'intelligibilité qu'il a lui-même construits à son niveau.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L., Cardinet, J. & Perrenoud, P. (Éd.). (1979). *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.
- Allal, L. & Laveault, D. (2009). Assessment for Learning. Évaluation-soutien d'apprentissage. *Mesure et évaluation en éducation*, 32(2), 99-100.
- Allal, L. & Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. In OCDE, *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires* (pp. 265-290). Paris : OCDE.
- Bachelard, G. (1966). *La formation de l'esprit scientifique* (2^e éd.). Paris : Vrin.
- Barbier, J.-M. (Éd.). (1996). *Savoirs théoriques et savoirs d'action*. Paris : Presses Universitaires de France.
- Beillerot, J. (2001, janvier). *Formation et Université*. Conférence donnée au 14^e colloque de l'ADMEE-Europe, Aix-en-Provence.
- Besnier, J.-M. (2005). *Les théories de la connaissance*. Paris : Presses Universitaires de France.
- Bloom, B. S., Hastings, J. T. & Madaus, C. F. (Éd.). (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York : Mac Graw Hill.
- Bonami, M. & Garant, M. (1996). *Systèmes scolaires et pilotage de l'innovation : émergence et implantation du changement*. Bruxelles : De Boeck.
- Bourdieu, P. & Passeron, J.-C. (1964). *Les héritiers*. Paris : Minuit.
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris : Presses Universitaires de France.
- Cardinet, J. (1977). *Objectifs pédagogiques et fonctions de l'évaluation*. Neuchâtel : INRP.
- Chevallard, Y. (1986). Vers une analyse des faits didactiques. In J.-M. De Ketele (Éd.), *L'évaluation : approche descriptive ou prescriptive ?* (pp. 31-67). Bruxelles : De Boeck.
- Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck.
- De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, 103, 59-80.
- De Ketele, J.-M. (2000). Approche socio-historique des compétences dans l'enseignement. In C. Bosman, F.-M. Gerard & X. Roegiers (Éd.), *Quel avenir pour les compétences ?* (pp. 83-92). Bruxelles : De Boeck.
- De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, 15(1), 25-37.
- De Ketele, J.-M. & Gerard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.
- De Ketele, J.-M. & Maroy, C. (2011). Quels critères de qualité pour la recherche en éducation ? In L. Paquay, M. Crahay & J.-M. De Ketele

- (Éd.), *L'analyse qualitative en éducation. Des pratiques de recherche aux critères de qualité* (pp. 223-253). Bruxelles : De Boeck.
- De Ketele, J.-M. & Paquay, L. (2011). Quels critères de qualité dans les recherches empiriques ? Le cas des recherches en éducation. In P. Servais (Éd.), *L'évaluation des recherches en sciences humaines et sociales. Regards de chercheurs* (pp. 247-256). Louvain-la-Neuve : Academia-Bruylant.
- De Landsheere, G. (1980). *Examens et évaluation continue. Précis de docimologie*. Bruxelles : Labor.
- Depover, C. & Noël, B. (Éd.). (1999). *L'évaluation des compétences et des processus cognitifs*. Bruxelles : De Boeck.
- Dupriez, V. & Vandenberghe, V. (2004). L'école en Communauté française de Belgique : de quelle inégalité parlons-nous ? *Les Cahiers du GIRSEF*, 27, 3-26.
- Duru-Bellat, M. (2002). *Les inégalités sociales à l'école*. Paris : Presses Universitaires de France.
- École supérieure de l'éducation nationale, de l'enseignement supérieur et de la recherche. (2009). *Épistémologie et méthodologie de l'évaluation*. Poitiers : ESEN.
- Figari, G. (1994). *Évaluation : quel référentiel ?* Bruxelles : De Boeck.
- Figari, G. (2006, septembre). *Pour une épistémologie de l'évaluation*. Conférence donnée au 19^e colloque de l'ADMEE-Europe, Luxembourg.
- Gilly, M. (1980). *Maître-élève. Rôles institutionnels et représentations*. Paris : Presses Universitaires de France.
- Jorro, A. (2006). Devenir un ami critique. Avec quelles compétences et quels gestes professionnels ? *Mesure et évaluation en éducation*, 29(1), 31-44.
- Jorro, A. (Éd.). (2009). *La reconnaissance professionnelle en éducation ? Évaluer, valoriser et légitimer*. Paris : L'Harmattan.
- Kuhn, T. S. (1972). *La structure des révolutions scientifiques*. Paris : Flammarion.
- Lalande, A. (1968). *Vocabulaire technique et critique de la philosophie*. Paris : Presses Universitaires de France.
- Le Moigne, J.-L. (1999). *La modélisation des systèmes complexes*. Paris : Dunod.
- Lieberman, A. (Éd.). (2005). *The roots of educational change. International handbook of educational change*. Dordrecht : Springer.
- Maroy, C. (2000). *L'enseignement secondaire et ses enseignants*. Bruxelles : De Boeck.
- Maroy, C. (2007). L'école à la lumière des sociologies des organisations. *Les Cahiers du GIRSEF*, 49.
- Martinand, J.-L. (1986). *Connaître et transformer la matière*. Berne : Peter Lang.
- Noizet, G. & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.

- Perrenoud, P. (1984). *La fabrication de l'excellence scolaire dans l'enseignement primaire : du curriculum aux pratiques d'évaluation. Vers une analyse de la réussite, de l'échec et des inégalités comme réalités construites par le système scolaire.* Genève : Droz.
- Piaget, J. (1970). *Épistémologie et sciences de l'homme.* Paris : Gallimard.
- Piéron, H. (1963). *Examens et docimologie.* Paris : Presses Universitaires de France.
- Popper, K. R. (1982). *La connaissance objective.* Paris : Éditions complexes.
- Rebber, B. (2010). *Théories politiques, pluralisme éthique et épistémologie de l'évaluation des technologies controversées.* Thèse HDR en Lettres et Philosophie, Université Paris Descartes.
- Ricœur, P. (2004). *Parcours de la reconnaissance.* Paris : Seuil.
- Scriven, M. (1967). *The Methodology of Evaluation.* Chicago : Rand Mac Nally.
- Stobart, G. (2011). L'évaluation pour les apprentissages : d'une expérimentation locale à une politique nationale. *Revue française de pédagogie*, 174(1), 41-48.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O. & Provus, M. M. (1980). *L'évaluation et la prise de décision en éducation.* Victoriaville : N.H.P.
- Vandenberghe, V. (1999). Combining Market and Bureaucratic Control in Education. *International Comparative Education*, 35(3), 271-282.
- Weiss, J. (1984). *La subjectivité blanchie ?* (IRD/R 84.17b). Neuchâtel : IRDP.

Configurations pour une épistémologie de l'évaluation

**Gérard Figari
Université de Grenoble**

Oser traiter le sujet de l'épistémologie de l'évaluation ? Pour répondre à une telle question, commençons par poser la question contraire : pourquoi n'en parlerait-on pas ? Si l'on accepte de postuler que, premièrement, l'évaluation englobe des phénomènes suffisamment prégnants dans le fonctionnement des dispositifs d'éducation et dans la réussite des apprentissages pour poser d'innombrables questions qui appellent des réponses ; que, deuxièmement, ces réponses doivent passer par des investigations rigoureuses ; que, troisièmement, les résultats espérés demanderont à être validés par la confrontation avec des modèles reconnus ou des modélisations argumentées, alors intervient naturellement la nécessaire élaboration d'une réflexion épistémologique sur l'activité évaluative dans son ensemble. À charge, pour les débatteurs, d'en construire le schéma organisateur, en utilisant l'opportunité d'une rencontre d'auteurs dont les textes ont été sollicités sur des thèmes connexes (savoirs produits autour de l'évaluation, modèles et modélisations de l'évaluation). C'est dans cet esprit qu'il a été proposé l'organisation d'une discussion sur les caractéristiques scientifiques de l'évaluation en éducation, avec l'intention de glisser du terrain de *la doxa* (niveau habituel des échanges sur ce sujet, que ce soit à l'école, dans les organismes de formation ou même à l'Université) vers celui de *l'épistémè* (domaine habituellement réservé aux épistémologues mais ouvert également aux chercheurs soucieux de situer leurs interrogations dans des problématiques de rapport au savoir, de modélisation, de référence à des « paradigmes »). Nous avons fait état de la « crise du sens » qui traversait l'ensemble du champ de l'évaluation et nous avons proposé aux coauteurs de cet ouvrage de partager avec nous la quête de modèles d'intelligibilité

utilisables par les évaluateurs, en les examinant sous l'angle épistémologique. Il s'agissait de convoquer ces modèles, de tenter de les rapprocher, pour avancer vers des dialogiques stables, sinon des consensus, pouvant représenter les (ou des) axes d'un « paradigme », un « modèle », un « mode de pensée » émergent, reconnaissable comme représentant des conceptions de l'évaluation dominant aujourd'hui le domaine de l'éducation. L'expérience de cet ouvrage a donc consisté à explorer les ouvertures à une réflexion sur les liens entre évaluation et démarche scientifique, à travers l'examen des « modélisations » sur lesquelles s'appuient, se fondent ou se construisent les procédures qui la mettent en œuvre. À partir des écrits qui précèdent, on peut, à présent, dessiner une configuration des réponses d'ordre épistémologique formulées autour des questions posées aux contributeurs, portant sur les savoirs produits et les modélisations invoquées par l'évaluation en éducation. C'est donc, tout d'abord, le lieu de faire l'inventaire des positions, des oppositions et des propositions avant de mesurer le cheminement d'une théorisation : celle consistant à identifier et à schématiser les dimensions épistémologiques de l'évaluation en éducation, telle qu'elle est *pensée* et non pas seulement *pratiquée* aujourd'hui.

Nous avons voulu rassembler et interpréter les prises de position épistémologiques des auteurs de l'ouvrage autour de trois dimensions : 1) les liens entre évaluation et pratiques scientifiques ; 2) les apports de la modélisation ; 3) la formulation de « paradigmes » en guise d'épistémologie. Tout en s'arrangeant avec les divergences et les convergences pour en développer des configurations possibles.

ÉVALUATION ET PRATIQUES SCIENTIFIQUES

En préalable, des fondements théoriques sous-jacents aux pratiques d'évaluation

Nous avons déjà noté que la littérature offre des avis partagés sur le fait de savoir si les processus de l'évaluation relèvent de fondements théoriques. Mais, dans le présent ouvrage, il apparaît, néanmoins, que le recours à des savoirs de référence, des modèles ou des modélisations, constitue la tendance majoritaire. Valois, Houssemand et de Leeuw adoptent cette position : il s'agit d'élaborer le « cadre théorique » censé définir ce que l'on « essaie de mesurer » avant le « cadre empirique » qui « précisera comment on va le mesurer »¹. Lafontaine et Monseur renforcent l'idée, à propos de « l'élaboration et de la mise en œuvre des enquêtes internationales » (PISA), en posant

1. Les nombreuses citations des chapitres précédents sont pour la plupart des citations aménagées en vue de la concision et de la fluidité du propos. C'est pourquoi elles ne comportent aucun renvoi.

la mobilisation préalable de « deux types de théories : celles relatives au domaine évalué et celles relatives à la mesure ». Les travaux décrivant des procédures de modélisation, dans la seconde partie de l'ouvrage, construisent également une grille de lecture théorique (interactionnisme, mesure du comportement, théories de la régulation, systémique, politique) avant de passer aux élaborations schématiques, explicatives et interprétatives. Enfin, même si Allal tient à séparer, d'entrée de jeu, objets d'évaluation et cadres théoriques, cette auteure appuie une présentation des paradigmes de référence utilisés dans l'évaluation des dispositifs de formation sur des axes théoriques majeurs (« post-positivistes, constructivistes, transformatifs et pragmatiques »).

Des prises de position épistémologiques

Constatons, tout d'abord, que plusieurs acceptions de l'épistémologie sont implicitement concernées dans ce recueil, même s'il n'est pas prévu d'ouvrir ici un débat sémantique, scientifique et philosophique sur ce concept majeur. Nous nous limiterons à relever que deux conceptions au moins sont en présence : une approche classique de l'épistémologie des sciences (selon Bachelard [1971] ou Popper [trad. 1991], par exemple), étudiant les conditions de production du savoir scientifique, et une approche plus générale, considérant « comment les êtres humains réfléchissent et pensent » (selon Fourez [2003] p. 9), et préconisant une réflexivité et une mise en doute systématique par rapport à l'activité éducative ou évaluative. La coexistence de ces deux niveaux de réflexion ne facilite pas la formulation des termes d'un éventuel débat entre des positions alternatives clairement identifiables ; toutefois, puisque les deux sortes d'éclairage enrichissent ce qui peut se définir comme une « pensée sur l'évaluation », nous les retiendrons conjointement.

Dans cet ouvrage, comme dans l'ensemble de la littérature, on trouve, sur l'évaluation, des positions tantôt « a-épistémologiques » (qui situent les problématiques d'évaluation dans le périmètre strict de la pratique et de ses méthodes, au service de la prise de décision et du jugement), tantôt « pré-épistémologiques » (qui introduisent, dans une approche réflexive, la référence à des modèles d'intelligibilité gouvernant les choix méthodologiques), tantôt ouvertement « épistémologiques » (qui interrogent les pratiques en termes de dimension scientifique, de modélisation et de production de connaissances et/ou de rapport au savoir). La lecture consécutive de l'ensemble des contributions autour des axes d'interrogation des éditeurs laisse apparaître, en même temps que des ancrages différenciés et des options scientifiques préalables souvent marquées, des points de débat ouverts, des rapprochements insoupçonnés, et des esquisses de convergences en matière de théorisations de l'évaluation qu'il nous appartient de décrypter ici.

La question de la dimension scientifique des pratiques d'évaluation

On va retrouver, chez les auteurs, les signes d'une sensibilité à l'examen préalable de l'éventuelle dimension scientifique de l'évaluation, à travers quatre thématiques, au moins : la prise de position distanciée vis-à-vis du positivisme, la distinction entre pratique et recherche, le rapport au savoir et le rôle des acteurs-participants dans la construction du sens.

DES CRITIQUES DE L'ASSUJETTISSEMENT DE L'ÉVALUATION AU POSITIVISME SYSTÉMATIQUE

Alors que nous engageons cette discussion dans le contexte d'une histoire de l'évaluation en éducation qui a fait la part belle à la mesure et que la doxa ambiante reste proche de ces influences originelles, la critique des approches « positivistes », qui furent celles des travaux fondateurs de l'évaluation en éducation, constitue un point faisant consensus parmi les contributeurs ayant traité ce sujet. On trouve, ainsi, chez Rodrigues et Machado, le rappel de la critique déjà ancienne faite à « l'opposition fondatrice de la dichotomie entre évaluateur et évalué, sujet et objet, etc. » selon laquelle « l'activité d'évaluer serait un processus objectif de mesure de la performance » au profit exclusif du sujet évaluateur. Cette « remise en cause épistémologique et sociale » a constitué une étape importante de la redéfinition du processus évaluatif en direction des conceptions annonciatrices de l'apparition des concepts d'évaluation formative (Allal, Cardinet & Perrenoud, 1979) et, plus tard, d'évaluation pour l'apprentissage (voir Stiggins, 2005, cité par Laveault dans le présent ouvrage). Dans le prolongement de cette critique apparaît celle, très générale aussi, dénonçant les incertitudes de la notation (Merle). Cet auteur reprend les conclusions des études docimologiques qui ont démontré « l'utopie » constituée par la notation ne constituant qu'une mesure « approximative » de la compétence des élèves. Mais la critique englobe aussi l'approche « déterministe », au sens large, qui mesure les effets des « variables contextuelles » sur les résultats obtenus, ce qui engendre de nombreux « biais » sur les jugements évaluatifs. De Ketele qualifie, enfin, « l'évaluation mesure » de réductrice et de faussement scientifique du fait qu'elle ne rend pas compte des « tâches complexes mutualisées » et que « la volonté d'objectivation » qui la caractérise conduit à « sélectionner les items les plus facilement évaluables ». Ce n'est donc pas parce qu'une évaluation ne serait pas fondée sur un modèle déterministe et formellement quantitatif, qu'elle n'obtiendrait pas son statut scientifique : dont acte.

UNE DISTINCTION « ÉVIDENTE » ENTRE PRATIQUE ÉVALUATIVE ET RECHERCHE SUR L'ÉVALUATION ?

La distinction habituelle entre pratique évaluative et recherche sur l'évaluation semble faire consensus chez les contributeurs à cet ouvrage. Sont ainsi généralement distinguées :

- Les pratiques d'évaluation, actes « d'expertise spécialisée » (Allal) à rapprocher des « actes sociaux » de Barbier (2001) produisant des résultats indicatifs, provisoires, et utilisables principalement par les acteurs (élèves/apprenants, enseignants, parents, décideurs) ; les savoirs (ou connaissances) élaborés par ces pratiques seraient alors locaux, contextualisés, souvent uniques et non reproductibles.
- Les recherches sur les phénomènes liés à l'évaluation, qui produisent des résultats critiquables par la communauté des chercheurs. Ils seraient donc d'une portée plus générale, transposables et généralisables, selon les règles fixées par les méthodes des sciences humaines et sociales.

Mais cette opposition est tellement simple et d'apparence tellement évidente qu'elle suscite un doute. En effet, cette doxa bien établie, au nom de laquelle il est coutume de séparer l'activité évaluative et la recherche sur cette activité, satisfait à la fois les praticiens, les décideurs et les chercheurs. En effet, cette position entraîne un confort total, aussi bien chez les évaluateurs ainsi invités à instrumenter, sans justifications scientifiques précises, que chez les décideurs ainsi légitimés dans leurs croyances et leur marge d'arbitraire, de même que chez les chercheurs, enfin, ainsi confortés dans la validité de leurs modèles de référence pour produire des savoirs « sur » l'évaluation, considérée, uniquement, comme objet d'étude de leur propre discipline. Et, pourtant, l'évaluation elle-même est, à la fois, pratique et investigation : avant de décider ou de juger, elle procède à une quête d'informations, elle-même guidée par des questions, des problématiques, des méthodologies de traitement de données, des interprétations. Est-il toujours possible de déterminer si un extrait textuel traitant de l'analyse d'une situation d'enseignement/apprentissage appartient à un compte rendu de recherche ou à un rapport d'évaluation ? Dans le même sens, de nombreux travaux récents (sur l'évaluation de l'expérience et des compétences, sur l'évaluation située, sur l'évaluation pour l'apprentissage, sur l'évaluation interactive, par exemple) mettent en avant la complexité du processus évaluatif, profondément impliqué dans une dialectique entre théorie et pratique : c'est aussi le cas, d'ailleurs, pour un certain nombre de contributions de cet ouvrage. Nous avons, quant à nous, postulé (Figari, 2006) qu'il existait un lien entre connaissances locales et savoirs généraux produits par les différentes activités d'évaluation (pratiques utilisant une méthodologie rigoureuse, d'une part, et recherche d'autre part). On rencontre, en effet, dans certaines situations, comme les études nationales ou internationales sur les compétences des élèves (Lafontaine & Monseur), une complémentarité sensible entre ces deux types de production, qui donne à penser que l'activité évaluative, si elle ne peut prétendre être scientifique, comporte, en revanche, une dimension heuristique renouvelable en permanence : chaque résultat, en étant réinvesti dans le contexte qui l'a produit, provoque un nouveau questionnement qui fait évoluer le précédent. On

retrouve, très clairement formulée, la même conception chez Morrisette, Mottier Lopez et Tessaro qui concluent que de nouveaux cadres théoriques apparaissent dans la confrontation entre acteurs, qui « insufflent un sens nouveau au matériau » déjà recueilli. L'évaluation désignerait donc une « réalité » plurielle, un « continuum » entre traitement d'informations situées et recherche (ce qui, évidemment, s'inscrirait contre une approche bachelardienne de la discontinuité). Allons un peu plus loin : si l'on reconnaît, dans toute pratique d'évaluation réfléchie, fondée et instrumentée, la possible existence d'une question, d'une hypothèse, d'une méthodologie de traitement de données empiriques et de résultats analysés, il reste peu de raisons pour ne pas s'interroger sur le type de connaissances produit, surtout s'il se réfère à un modèle théorique qui en oriente le sens en même temps que peut se vérifier la validité des méthodes employées.

La question de la relation de l'évaluation au savoir : quel(s) savoir(s) ?

SAVOIR OU CONNAISSANCE ?

La relation qu'entretient l'activité évaluative avec la production de savoir reste peu étudiée, en tant que telle, et lorsqu'elle est traitée, comme c'est le cas dans ce recueil, elle affleure dans des problématiques diverses ne présentant que peu de liens entre elles. Certaines de ces problématiques examinent le statut des *connaissances* issues de l'évaluation alors que d'autres cherchent à formuler les types de *savoirs* qu'elle peut contribuer à produire. On retrouve, ici, la distinction parfois opérée entre savoir et connaissance, ne serait-ce qu'au niveau des termes employés : d'un côté, des données répertoriées et reconnues (constituant des champs de *savoir* sur l'évaluation) et de l'autre, un sujet construisant ses propres *connaissances* (éventuellement à travers la transaction évaluative). Cette distinction (bien que discutée) peut déjà constituer un reflet de la position des auteurs au sujet de ce questionnement épistémologique : la pratique évaluative participerait au processus de construction d'une connaissance chez l'enseignant comme chez l'apprenant alors que, seule, la recherche sur l'évaluation pourrait prétendre produire des savoirs réutilisables.

ÉVALUATION, PRODUCTION DE SAVOIRS, CONSTRUCTION DE CONNAISSANCES

Sur cette thématique, même si les points de vue des auteurs paraissent diversifiés et peu consensuels, quelques postulats émergent, confortant une position relativement ouverte : si des *savoirs* pouvaient être produits, ils seraient de l'ordre du social ou du stratégique, et si des *connaissances* étaient en jeu, elles ne seraient que référentielles (« des connaissances de référence », selon De Ketele).

- A propos d'évaluation comme *production de savoirs*, on trouvera, en effet, surtout, l'évocation de « savoirs pratiques » créés par l'évaluation formative, « objectivés par des négociations de sens » (Morrisette *et al.*), de « savoirs explicatifs ou normatifs » produits à l'occasion d'enquêtes internationales (Lafontaine & Monseur), de « savoirs-jugements » ou de « savoirs d'action », produits par des processus « ayant pour finalité de prendre des décisions pour l'action » (De Ketele). Ces deux derniers avatars du « savoir » évaluatif (savoir-jugement et savoir ayant pour finalité de prendre des décisions) donnent la mesure du statut pouvant être attribué à celui-ci, dans les résultats de l'évaluation : le jugement et l'action ne sont que des appositions au savoir mais jamais le savoir lui-même. Quant aux « savoirs sur l'évaluation » produits par la recherche, ils sont validés par les communautés scientifiques qui les contrôlent, puisqu'ils appartiennent au champ des sciences sociales.
- Les liens de l'évaluation avec la *construction de connaissances* sont traités selon la même approche. Tout d'abord, De Ketele distingue, comme nous, des « connaissances globales, appelées aussi standards » et des « connaissances locales ». Mais l'auteur restreint ensuite le sens de ces connaissances en les assimilant à des « référents », en opposition à des « connaissances produites » par la recherche. Tourmen et Mayen, tout en évoquant la même idée, semblent la nuancer, en spécifiant que les connaissances issues de l'expérience « deviennent » des référents lorsqu'elles sont « utilisées comme éléments de comparaison pour attribuer un sens aux phénomènes observés ». Ce qui paraît sensible, ici, c'est une tendance à circonscrire la capacité de l'évaluation à contribuer à la production de savoirs à une simple contextualisation facilitant l'attribution de sens à un ensemble de données : savoir d'action, savoir local, savoir de référence.

LA PARTICIPATION DES ACTEURS DE TERRAIN À L'ÉLABORATION DE SAVOIRS

Dernière source de production de savoir proposée, plus spécifique, et justifiant peut-être une place à part : la participation des acteurs, abordée sous divers aspects. Tout d'abord, à travers les approches collaboratives : selon Morrisette *et al.*, ces dernières peuvent montrer que « la participation des acteurs est requise dans l'élaboration des savoirs produits par et sur l'évaluation ». Ensuite, par la médiation des participants eux-mêmes à l'évaluation d'une formation : Rodrigues et Machado instituent une typologie convaincante de « logiques de participation des acteurs » à l'évaluation. Ce faisant, ils établissent l'« intégration du processus de participation à l'évaluation » qu'ils modélisent, de façon à créer une grille de traitement des données en relation avec cette dimension qui entraîne une production de résultats de type social et politique. Sans que le savoir ainsi constitué soit défini ou étiqueté, on voit bien, ici, qu'il ne pourra qu'être un savoir

collectif et partagé. Enfin, Vial rappelle, en citant Feldman et Le Grand (1996), « l'existence et l'importance d'un savoir profane », doté d'une valeur sociale, apparenté, selon nous, à ceux de l'action et de l'expérience, déjà évoqués. Que ce concept puisse englober les précédents ou qu'il les rejoigne dans un faisceau de savoirs coconstruits par les acteurs (tendance émergente dans les productions évaluatives), nous apercevons ici une occasion pour d'éventuelles convergences.

Le « village des évaluateurs »

Nous avons pris note des « savoirs pratiques » développés dans les évaluations formatives, des « connaissances issues de l'expérience » décrites dans les travaux inspirés par la didactique professionnelle, des « savoirs explicatifs ou descriptifs » émanant des enquêtes internationales et des « savoirs-jugements » évoqués plus généralement en sciences de l'éducation. On voit bien, là, qu'il s'agit, essentiellement, de différentes formes d'un savoir non académique qui relèverait davantage du « profane », seul prétendant à figurer dans un questionnement épistémologique de l'activité évaluative. Cette dernière configuration de savoirs baliserait-elle un territoire légitime de l'activité scientifique relative à l'évaluation ? Il existerait, donc, pour abriter cet ensemble de pratiques, de problématiques, de recherches dédiées, d'entre-deux réunissant plusieurs de ces activités, ce que l'on pourrait appeler, par métaphore, « le village des évaluateurs », celui émancipé des autres territoires scientifiques dont il est partiellement issu (ceux de la sociologie ou de l'ethnologie, ceux des psychologies ou des sciences cognitives, ceux de la philosophie ou de l'anthropologie) mais qui se cherche encore un « référentiel commun ». Un lieu de convergences, de production ou d'explicitation de savoirs divers et encore émiétés, concernant les processus d'évaluation en situation éducative, lieu qui peine encore à définir, d'une même voix, son identité. Un village connu, répertorié... mais dont le nom et le contenu restent encore à valider.

MODÈLES ET MODÉLISATIONS AU SERVICE DE L'ÉVALUATION

Des opportunités de configurations plutôt qu'un éclatement de « modèles » ou de modélisations

La deuxième problématique qui émerge de la mise en perspective des textes présentés est, évidemment, celle, sollicitée, des *modèles* de l'évaluation. Nous étions partis d'une interrogation sur le recours à des modèles

pour justifier les choix évaluatifs, en tentant de définir ce que ce concept recouvre chez différents auteurs. Et, dans le contexte de l'évolution centrifuge des sciences sociales en général et des sciences de l'éducation en particulier, nous avons constaté, également, un éclatement des modèles potentiellement mobilisables dans le domaine de l'évaluation en éducation. Les textes rassemblés ici prolongent et nuancent en même temps cette tendance. Sont évoqués, successivement, des *modèles conceptuels* de l'évaluation (comme, par exemple, celui de « régulation interactive », cité par Morrissette *et al.*, permettant de « structurer » l'activité d'évaluation), des *modèles explicatifs* des phénomènes relevant de l'évaluation (comme, par exemple, le modèle interactionniste, présenté par Merle), des *modèles désignés* comme *référentiels sous-jacents* par Demailly, assez proches de ceux pouvant s'appeler « *paradigmes* », selon De Ketele, des *modèles prescriptifs* de « pratiques efficaces » (formulés par Laveault). L'éventail reste très large et le lecteur a l'impression d'accéder à des discours divers sur le recours à la notion de modèle dans les choix méthodologiques inhérents aux activités d'évaluation.

On peut également dégager des textes présentés deux sortes de postures face au recours potentiel à des modèles d'évaluation : une posture passive et une posture active. Par commodité, on appellera posture passive celle qui consiste à mobiliser un modèle connu, validé dans un champ scientifique ou disciplinaire précis, comme grille de lecture du phénomène étudié. Les exemples les plus clairs figurent chez Merle et chez Valois *et al.* Merle utilise le modèle interactionniste pour « mieux rendre compte de la diversité des situations scolaires appréhendées » et pour « rendre intelligible la réussite scolaire en milieu populaire ». Quant à Valois *et al.*, ils opérationnalisent un modèle d'attitude illustrant « la théorie du comportement planifié », opération à partir de laquelle peuvent alors s'organiser les différentes phases de la mesure évaluative. À la différence de cette posture dite « passive » face à un modèle (le modèle existe, on ne le crée pas), la posture dite « active » représente donc l'ensemble des démarches entreprises pour construire ou reconstruire un modèle apte à rendre compte d'une situation d'évaluation particulière. Ainsi, selon Vial, modéliser est « un concept souvent utilisé pour désigner une formalisation généralisante » ; « c'est un processus, non une procédure ». Selon Morrissette *et al.*, il s'agit d'élaborer « une représentation schématique devenue un artefact aménageable [...] constituant une grille théorique ». Quant à Rodrigues et Machado, ils conçoivent que modéliser consiste à « construire des référentiels avec l'expérience et avec la réflexion que le processus d'évaluation même suscite ». Ces prises de position amènent à concevoir l'hypothèse d'une sorte de consensus sur l'intérêt de disposer, dans le déroulement de l'activité évaluative, d'outils conceptuels intermédiaires entre des « paradigmes » et des critères, aptes à expliquer les résultats de différentes opérations éparses grâce à un comparant unique et reconnaissable : ces outils seraient appelés « modèles ».

Les études et les essais qui précèdent auraient contribué à tracer quelques dénominateurs communs aux différentes façons de mobiliser des modèles.

Une configuration de niveaux de modèles d'évaluation reconnaissables ?

On peut tenter de concevoir une telle configuration de modèles d'évaluation autour des trois niveaux que les textes de ce recueil semblent avoir dégagés : le niveau du *modèle conceptuel* (rattaché à une théorie : du fonctionnement des systèmes, de la mesure, de l'apprentissage, de l'activité, etc.) ; le niveau du *modèle explicatif*, sorte de grille de lecture évaluative utilisable dans des situations semblables (évaluation d'apprentissages précis, évaluation comparée de dispositifs éducatifs, par exemple) ; le niveau du *modèle prescriptif*, de « pratique efficace » (formatif, métacognitif, etc.). Dans les trois figures, on retrouve le rôle affecté au modèle par Machado, rôle qui permet de construire « un ensemble d'hypothèses de travail à partir desquelles peut se développer une consolidation empirique »² (2007, p. 66). Le modèle assume, ainsi, un rôle heuristique permettant à l'évaluateur d'accéder à l'intelligibilité du réel en en prélevant des éléments que la modélisation (active ou passive) a désignés comme caractéristiques. Nous postulons que, dans ces trois niveaux de modélisation, les modèles d'intelligibilité du réel proposés par les auteurs sont reconnaissables et échangeables dans la communauté des évaluateurs (praticiens et chercheurs), même au-delà du « village » qui les réunit.

DES PARADIGMES D'ÉVALUATION EN GUISE D'ÉPISTÉMOLOGIE ?

Le troisième terme du débat épistémologique ayant émergé dans une partie notable des textes présentés est constitué par les *paradigmes de l'évaluation*.

L'invocation de « paradigmes d'évaluation »

Dans la quête d'intelligibilité pour interpréter les résultats de l'évaluation, le recours au concept de « paradigme » se répand dans le « village des évaluateurs », au moins depuis la publication de De Ketele (1993). Par la suite, plusieurs utilisations du terme « paradigme d'évaluation » ont été formulées par des auteurs ; elles se retrouvent dans différents textes présentés ici, principalement autour de trois catégories : la référence partagée entre

2. C'est nous qui traduisons.

chercheurs d'une même communauté scientifique, une sorte de vision du monde, un ensemble de modèles d'évaluation.

LE PARADIGME SCIENTIFIQUE

Une première façon de parler des paradigmes dans le discours évaluatif consiste, tout naturellement, à utiliser l'acception scientifique du terme, souvent référée à Kuhn (1970), comme précisé dans l'introduction. Elle est également invoquée, de manière générale, par Allal qui rappelle ici le sens de « référence partagée » que le concept prend souvent dans les recherches évaluatives. Cette acception est également présente chez Merle, sous la forme du recours au « paradigme de l'individu » et à celui du « contexte historique », en sociologie. Ayant fait appel à des modèles explicatifs de la notation scolaire, le même auteur invoque, tour à tour, les paradigmes déterministe et interactionniste comme façons de traiter les phénomènes qui en relèvent. Enfin, Lafontaine et Monseur limitent l'emploi du concept à la démarche de recherche (« paradigme de recherche », « paradigme positiviste »).

LE PARADIGME COMME VISION DU MONDE

Dans le prolongement de la première acception du terme, Rodrigues et Machado utilisent le concept de paradigme comme la formalisation d'une vision du monde, une construction métathéorique qui vise à organiser et à systématiser les conceptions et les pratiques de l'évaluation. Les trois paradigmes proposés par ces auteurs prétendant regrouper et classer les différentes postures d'évaluation possibles en *objectivistes*, *subjectivistes* et *interactives-dialectiques*, rejoignent de manière assez fidèle la typologie de Cardinet (*évaluation externe, interne et négociée*) et peuvent aussi, dans une moindre mesure, se rapprocher, si on les regroupe, de ceux formulés par De Ketele (*évaluation-mesure, évaluation intuitive-pragmatique, évaluation des/pour les apprentissages*). Vial, quant à lui, n'a pas recours au concept de « paradigme », mais on peut se demander si l'expression de « mode de pensée » qu'il utilise n'en constitue pas une autre facette qui semble même se prolonger à travers le sens qu'il attribue au concept de « modèle d'évaluation », défini comme « un ensemble organisé de principes d'action, dépendant, chez le sujet évaluateur, de théories ». Cette approche du paradigme comme vision du monde, tout en respectant le sens originel du concept, l'élargit de telle façon qu'il devient apte à définir la subordination de l'action évaluative à des finalités identifiées, consacrant son statut téléologique, voire philosophique, sinon scientifique.

LE PARADIGME COMME ENSEMBLE DE MODÈLES D'ÉVALUATION

Même si De Ketele part de la définition canonique de paradigme comme « cadre à l'intention d'une communauté scientifique », il conserve le terme pour parler des modèles utilisés par les évaluateurs lorsqu'ils construisent la vision de l'activité qu'ils déploient. Le concept devient alors « un modèle

privilegié à suivre [...] à des époques ou dans des lieux donnés ». Quant à Merle, il utilise également la notion de paradigme dans un raisonnement méthodologique, lorsqu'il évoque, par exemple le « paradigme statistique » ou le « paradigme interprétatif », davantage applicables, en effet, au déroulement d'un protocole évaluatif. La coexistence de ces différentes formulations, dans les travaux exposés, nous amène à suggérer que, si le concept de paradigme, lui-même, appartient à la sphère de l'épistémologie et, si les auteurs l'appliquent à des « modèles » ou des « façons de penser » l'évaluation, c'est qu'il permet de dire *l'épistémè* (l'évaluation appartenant au questionnement scientifique et se réclamant de postulats, de règles et de méthodes qui engagent une communauté de chercheurs) tout en affirmant, en même temps, *la doxa* (l'évaluation demeurant une pratique établie de jugement, demandant seulement à être instrumentée).

Le recours aux « paradigmes » : un lien inéluctable entre recherche et évaluation ?

Si nous avons classé les réflexions des auteurs en matière de recours aux « paradigmes » pour décrire les raisonnements évaluatifs autour de quelques thématiques permettant des problématisations, il n'en reste pas moins que les textes sources restent plus complexes et plus nuancés que ce type de relevé peut le laisser voir. Les nuances et les interactions entre les thèmes abordés sont aussi contributives au débat que les assertions extraites de leur contexte. C'est pourquoi, ce relevé de formulations doit être complété par la remarque suivante : certains auteurs, en se réclamant de telle ou telle position, établissent un lien explicite ou implicite entre pratique évaluative et recherche scientifique, à travers ce recours aux « paradigmes ». Ainsi, on note qu'Allal, tout en rappelant Kuhn (1970), utilise le concept de paradigme, forgé dans le contexte de la recherche scientifique, pour « interroger les fondements épistémologiques de l'évaluation ». On peut se demander, alors, si l'évaluateur, à son niveau d'intervention, peut se débrouiller sans recours aux « paradigmes » pour référer ses protocoles et situer ses résultats, pour disposer d'une grille d'interprétation reliée à une vision plus générale de l'apprentissage, de ses processus comme de ses dispositifs. Qu'on le veuille ou non, qu'on le proclame ou non, les différents paradigmes définis, forgés ou invoqués par les contributeurs ne constituent-ils pas, chacun dans son contexte, le lien intrinsèque et inéluctable entre recherche et évaluation ?

POUR CONCLURE

On a observé la difficulté à dégager un sens homogène aux productions évaluatives en éducation, dans le contexte général de « pensée éclatée »

que semblent connaître les sciences humaines et sociales. On a noté la tendance de certaines disciplines à traiter le sujet de l'évaluation à partir de leurs paradigmes internes, accentuant ainsi la perte de légitimité de l'existence d'un champ de recherche spécifique à l'évaluation. Mais, contrairement à ce constat, une autre analyse, alimentée, en partie au moins, par les contributions présentées ici, argumente en faveur de l'existence possible d'un espace pour théoriser l'évaluation. Le cloisonnement habituel aux sciences sociales, exporté en sciences de l'éducation et, a fortiori, dans l'un de ses champs d'étude, l'évaluation, ne nous a pas interdit de questionner le caractère situé et par conséquent intégré des phénomènes que celle-ci recouvre. Les auteurs convoqués ont laissé voir, à la fois, une intransigeance sur les frontières entre évaluation et recherche scientifique et une ouverture sur la définition des connaissances que les pratiques d'évaluation contribuent à construire. Entre les lignes, nous avons cru trouver place pour l'amorce d'une réflexion, qui reste à développer, sur la pertinence éventuelle d'une « théorie pragmatique » de l'évaluation, en partie tributaire et en partie indépendante des débats scientifiques qui animent aujourd'hui la communauté des chercheurs en sciences humaines et sociales. Cette approche pragmatique s'entend aussi bien au sens de Perrenoud (1991) qui l'illustre à travers la fonction de « régulation par l'action et l'interaction » et la « confrontation avec le réel » (p. 60), qu'au sens de la théorie pragmatique des actes de langage (Austin, 1970/1991) qui met en évidence les « énoncés performatifs » dans l'usage du langage en situation, particulièrement prégnant dans les transactions évaluatives, même si cette dernière dimension n'a pas été explicitée dans ce recueil.

Le « village des évaluateurs » abritant tous les courants de pensée évoqués ici (mais aussi ailleurs) ainsi que toutes les potentialités de convergences, de théorisations unificatrices et de possibilités de co-construction d'un champ praxéologique spécifique, existe : cet ouvrage en est une actualisation partielle et provisoire. Il est habité d'une culture scientifique de l'évaluation qui transcende la tradition du « jugement » et de la « prise de décision » pour questionner la référence à des « paradigmes princeps » et repérer des formes de savoirs « sociaux », « stratégiques », « d'action », « de jugement », « profanes » que l'évaluation contribue à construire.

Quoi qu'il en soit, plutôt que de laisser, sans fin, en jachère, une épistémologie de l'évaluation non élucidée et, pensons-nous non suffisamment interrogée, les auteurs de cet ouvrage ont accepté de s'exprimer autour de la question : peut-on penser cette épistémologie ? Le débat initié au cours de l'élaboration de cet ouvrage et qui vient d'être donné à examiner ici est, naturellement, destiné à se poursuivre.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L., Cardinet, J. & Perrenoud, P. (Éd.). (1979). *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.
- Austin, J. (1991). *Quand dire c'est faire* (G. Lane, trad.). Paris : Seuil. (Original publié 1970).
- Bachelard, G. (1971). *La formation de l'esprit scientifique*. Paris : Vrin.
- Barbier, J.-M. (2001). Quelques questions pour la recherche dans le domaine de l'évaluation. In G. Figari & M. Achouche (Éd.), *L'activité évaluative réinterrogée : regards scolaires et socioprofessionnels* (pp. 351-360). Bruxelles : De Boeck.
- De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, 103, 59-80.
- Feldman, J. & Le Grand, J.-L. (1996). Savoirs savants, savoirs profanes. In J. Feldman, J.-C. Filloux, B.-P. Lécuyer, M. Slez & M. Vicente (Éd.), *Éthique, épistémologie et sciences de l'homme* (pp. 89-105). Paris : L'Harmattan.
- Figari, G. (2006, septembre). *Pour une épistémologie de l'évaluation*. Conférence donnée au 19^e colloque de l'ADMEE-Europe, Luxembourg.
- Fourez, G. (avec Larochelle, M.). (2003). *Apprivoiser l'épistémologie*. Bruxelles : De Boeck.
- Kuhn, T. S. (1970). *La structure des révolutions scientifiques*. Paris : Champs Flammarion.
- Machado, E. A. (2007). *Évaluation et participation. Une étude sur le rôle des acteurs dans l'évaluation de la formation continue*. Thèse de doctorat en Sciences de l'éducation, Universidade do Minho.
- Perrenoud, P. (1991). Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation*, 13(4), 49-81.
- Popper, K. R. (1991). *La connaissance objective* (J.-J. Rosaot, trad.). Paris : Aubier.
- Stiggins, R. J. (2005). *Assessment For Learning Defined*. Assessment Training Institute. Consulté le 18 janvier 2012 dans <http://www.docstoc.com/docs/14728977/ASSESSMENT-FOR-LEARNING-DEFINED>

Notices biographiques

Linda Allal est professeure honoraire (émérite) de l'Université de Genève. Après l'obtention de son doctorat en *Educational psychology* à *Michigan State University*, elle a exercé la fonction de professeur à la Faculté de psychologie et des sciences de l'éducation de Genève pendant 32 ans. Ses recherches ont porté sur les processus d'apprentissage et d'évaluation en situation scolaire, notamment dans le domaine de la production textuelle. Ses publications sur l'évaluation concernent les processus de régulation des apprentissages, en particulier par l'évaluation formative interactive et l'autoévaluation, ainsi que le jugement professionnel des enseignants.

Adresse : Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : linda.allal@unige.ch

Jean-Marie De Ketele est professeur émérite de l'Université catholique de Louvain et de la Chaire UNESCO en Sciences de l'Éducation qu'il a créée à Dakar en 1994. Sa carrière d'enseignant et de chercheur porte surtout sur les problématiques de l'observation et de l'évaluation. Il a été également actif en pédagogie universitaire. Il dirige deux collections aux Éditions De Boeck. Il a été distingué par le Conseil culturel mondial et nommé docteur *honoris causa* de l'Université de Dakar.

Adresse : Université catholique de Louvain
Faculté de Psychologie et des Sciences de l'Éducation
Place Cardinal Mercier, 10
B-1348 Louvain-la-Neuve
Courriel : jean-marie.deketele@uclouvain.be

Astrid de Leeuw est étudiante au doctorat en sciences de l'éducation et de la psychologie à l'Université du Luxembourg et chercheuse à l'unité de recherche EMACS (Centre de recherche sur la mesure en éducation et en sciences cognitives appliquées). Sa thèse de doctorat porte sur l'éducation au développement durable, et plus précisément sur les attitudes des enseignants et professeurs envers l'enseignement des comportements durables, et les attitudes des élèves et étudiants envers ces comportements.

Adresse : Université du Luxembourg
Faculté des Lettres, des Sciences Humaines,
des Arts et des Sciences de l'Éducation
Route de Diekirch
L-7220 Walferdange
Courriel : astrid.deleeuw@uni.lu

Lise Demailly est agrégée de philosophie et professeure de sociologie. Elle a été directrice adjointe de l'Institut Universitaire de Formation des Maîtres de Lille, chargée de la formation des formateurs. Elle est actuellement professeure émérite à l'Université de Lille 1 et membre du CLERSE (Centre lillois d'études et de recherches sociologiques et économiques). Ses travaux de recherche s'inscrivent en sociologie de l'éducation, de la santé mentale et, sur le plan théorique, en sociologie des professions, des institutions et des politiques publiques. Elle s'intéresse notamment à l'approche sociologique des métiers et activités professionnelles relationnelles et aux rapports entre connaissances rationnelles et action publique.

Adresse : Maison européenne des Sciences de l'Homme et de la Société
2, rue des Canoniers
F-59000 Lille
Courriel : lise.demailly@univ-lille1.fr

Gérard Figari est professeur émérite à l'Université de Grenoble où, comme directeur de recherche, il a fondé le Groupe de Recherche sur l'Évaluation et l'Ingénierie de la Formation (GREIF). Il a été président de l'ADMEE-Europe (Association pour le Développement des Méthodologies d'Évaluation en Éducation) de 2000 à 2004. Ses principaux domaines de recherche sont l'ingénierie du curriculum, l'évaluation des dispositifs éducatifs, l'évaluation des compétences dans le cadre de la professionnalisation (notamment : reconnaissance et validation des acquis de l'expérience), l'étude des référentiels.

Adresse : 468, route des Vières
F-38250 Villard de Lans
Courriel : gerard.figari@wanadoo.fr

Claude Houssemand est professeur à l'Université du Luxembourg et chercheur à l'unité de recherche EMACS (Centre de recherche sur la mesure en éducation et en sciences cognitives appliquées). Ses travaux de recherche se focalisent sur la mesure des compétences, le chômage et la transition école-emploi. Il est rédacteur européen de la revue *Mesure et évaluation en éducation*.

Adresse : Université du Luxembourg
Faculté des Lettres, des Sciences Humaines,
des Arts et des Sciences de l'Éducation
Route de Diekirch
L-7220 Walferdange
Courriel : claude.houssemand@uni.lu

Dominique Lafontaine est titulaire d'une licence en langues et littératures romanes et docteure en sciences de l'éducation. Elle est actuellement professeure à l'Université de Liège, où elle dirige le service d'Analyse des systèmes et des pratiques d'enseignement. Elle a contribué à la mise sur pied des évaluations externes en Communauté française de Belgique et a été responsable nationale de plusieurs études IEA et PISA. Depuis 1999, elle est membre du groupe international d'experts en lecture pour PISA. Ses travaux portent à la fois sur la lecture et sur la comparaison des systèmes éducatifs, en particulier l'équité.

Adresse : Université de Liège
Faculté de Psychologie et des Sciences de l'Éducation
Bâtiment B32
5, boulevard du Rectorat
B-4000 Liège
Courriel : dlafontaine@ulg.ac.be

Dany Laveaul est professeur titulaire à l'Université d'Ottawa. Il y a occupé la fonction de vice-doyen à la recherche de 1998 à 2002. De 1993 à 2001, il a été directeur de la revue *Mesure et évaluation en éducation*. En 1995, il recevait le prix de l'Association pour le développement des méthodologies de l'évaluation en éducation (Canada). En 2002, le ministère de l'Éducation de l'Ontario l'a nommé coprésident du comité sur l'enseignement de la lecture et depuis 2004, il agit en tant qu'expert à l'Office de la qualité et de la responsabilité en éducation. Ses recherches portent sur l'autoévaluation, la régulation des apprentissages et des systèmes éducatifs.

Adresse : Faculté d'éducation, Université d'Ottawa
145, rue Jean-Jacques-Lussier
Ottawa, Ontario K1N 6N5
Canada
Courriel : dlaveaul@uottawa.ca

Eusébio André Machado est professeur à l'Université Portucalense au Portugal et directeur du Doctorat en éducation. Il est chercheur aux Universités du Minho et d'Aveiro. Il a publié plusieurs ouvrages et articles sur la formation et le développement professionnel des enseignants, ainsi que sur l'évaluation des enseignants.

Adresse : Université Portucalense
541, rua Dr. António Bernardino Almeida
4200-072 Porto
Portugal
Courriel : eusebio@upt.pt

Patrick Mayen est professeur en sciences de l'éducation à AgroSup Dijon où il a dirigé l'Unité de Recherche Développement Professionnel et Formation jusqu'en 2012. Ses travaux de recherche se situent dans le courant de la didactique professionnelle et ils portent sur les conditions et les processus d'apprentissage et de développement dans et par le travail et dans et par la formation. Ses objets : formation par alternance, validation des acquis de l'expérience, formation et évaluation en situation de travail.

Adresse : AgroSup Dijon
26, boulevard Petitjean
B.P. 87999
F-21079 Dijon Cedex
Courriel : patrick.mayen@educagri.fr

Pierre Merle, agrégé de sciences sociales, est professeur de sociologie à l'Institut Universitaire de Formation des Maîtres de Bretagne et à l'Université européenne de Bretagne. Ses recherches portent essentiellement sur l'expérience subjective des élèves, notamment celle de l'humiliation en classe, les pratiques de notation des professeurs, la démocratisation de l'enseignement, les dynamiques de ségrégation scolaire dans les collèges et lycées français.

Adresse : IUFM de Bretagne
153, rue Saint-Malo
F-35043 Rennes Cedex
Courriel : pierre.merle@bretagne.iufm.fr

Christian Monseur est professeur à l'Université de Liège. Ses principaux domaines de recherche et d'enseignement portent sur les méthodologies quantitatives d'analyse des systèmes éducatifs et sur les enquêtes à large échelle telles que les études PISA de l'OCDE ou les études de l'IEA. Il a d'ailleurs travaillé quatre ans à l'*Australian Council of Educational Research* en tant que directeur adjoint et *data manager* de l'étude PISA 2000. Il est membre du *Technical Advisory Group* de l'étude PISA et a été membre du comité scientifique de plusieurs études de l'IEA.

Adresse : Université de Liège
Faculté de Psychologie et des Sciences de l'Éducation
Bâtiment B32
5, boulevard du Rectorat
B-4000 Liège
Courriel : cmonseur@ulg.ac.be

Joëlle Morrissette est professeure adjointe à l'Université de Montréal et chercheure associée au Centre de recherche interuniversitaire sur la formation et la profession enseignante (CRIFPE). Ses principaux domaines de recherche sont l'évaluation des apprentissages et les méthodologies qualitatives. Elle aborde ses objets d'étude sous un angle interactionniste et socioculturel.

Adresse : Pavillon Marie-Victorin, bureau C-514
Faculté des sciences de l'éducation, Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal, Québec H3C 3J7
Canada
Courriel : joelle.morrissette@umontreal.ca

Lucie Mottier Lopez est professeure associée à l'Université de Genève et responsable de l'équipe de recherche EReD (Évaluation, Régulation et Différenciation des apprentissages en situation scolaire et en formation). Ses principaux champs d'investigations portent sur les pratiques d'évaluation, avec un accent mis sur les processus de jugement-en-acte des enseignants, sur les pratiques d'évaluation collaborative et de modération sociale entre professionnels, et sur les processus de régulation située en classe.

Adresse : Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : lucie.mottier@unige.ch

Pedro Rodrigues est professeur à l'Institut d'Éducation de l'Université de Lisbonne au Portugal (depuis la scission de la Faculté de Psychologie et des Sciences de l'Éducation de cette université, où il a enseigné auparavant). Il est membre de l'Unité de Recherche et Développement en Éducation et Formation (UIDEF) de cet institut et a publié plusieurs travaux sur la méthodologie de l'évaluation (de programmes, de la formation et des enseignants) et ses fondements.

Adresse : Instituto de Educação da Universidade de Lisboa
Alameda da Universidade

1649-013 Lisboa
Portugal
Courriel : pedro.rodrigues@ie.ul.pt

Walther Tessaro est docteur en sciences de l'éducation, chargé d'enseignement à la Faculté de psychologie et des sciences de l'éducation de l'Université de Genève. Membre de l'équipe EReD (Évaluation, Régulation et Différenciation des apprentissages en situation scolaire et en formation), il s'intéresse aux processus d'enseignement et d'apprentissage en situation scolaire et de formation, et plus spécifiquement aux pratiques d'évaluation et de régulation différenciée des apprentissages. Avant d'assumer ses fonctions actuelles, il a été responsable des épreuves externes et formateur au Service de l'Évaluation de l'Enseignement primaire de Genève.

Adresse : Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : walther.tessaro@unige.ch

Claire Tourmen est maître de conférences en sciences de l'éducation à AgroSup Dijon. Elle a réalisé une thèse sur « Les compétences des évaluateurs, le cas des évaluateurs de politiques publiques », en contrat CIFRE (Conventions Industrielles de Formation par la Recherche) avec l'entreprise Euréval (Lyon). Elle a depuis élargi ses recherches à l'évaluation de l'expérience et des compétences, et à l'analyse d'activités de travail comme celles des dirigeants, des formateurs, des doctorants.

Adresse : AgroSup Dijon
26, boulevard Petitjean
B.P. 87999
F-21079 Dijon Cedex
Courriel : claire.tourmen@educagri.fr

Pierre Valois est professeur à l'Université du Luxembourg et chercheur à l'unité de recherche EMACS (Centre de recherche sur la mesure en éducation et en sciences cognitives appliquées). Ses intérêts de recherche portent sur la relation attitude-comportement, les théories de la mesure, l'évaluation des apprentissages et l'éducation à la santé. Il est directeur de la revue *Mesure et évaluation en éducation*.

Adresse : Université du Luxembourg
Faculté des Lettres, des Sciences Humaines,
des Arts et des Sciences de l'Éducation
Route de Diekirch
L-7220 Walferdange
Courriel : pierre.valois@uni.lu

Michel Vial est habilité à diriger des recherches, spécialiste de l'évaluation des relations humaines dans la relation éducative, l'évaluation des pratiques, l'évaluation située en éducation. Il est responsable d'une licence de formateurs responsables d'évaluation et d'un master de consultants et de coaches, accompagnateurs du changement en ressources humaines. Il anime un groupe de recherche sur l'accompagnement professionnel et la professionnalisation. Consultant et superviseur de professionnels, il intervient occasionnellement dans la fonction publique et territoriale, les métiers de la santé sociale et les entreprises.

Adresse : Université de Provence
Sciences de l'éducation
29, avenue Robert-Schuman
F-13100 Aix-en-Provence
Courriel : michel.vial@univ-amu.fr

Table des matières

Sommaire	5
Modèles et modélisations face à une crise du sens de l'évaluation en éducation	7
<i>Lucie Mottier Lopez* & Gérard Figari**</i>	
<i>*Université de Genève</i>	
<i>**Université de Grenoble</i>	
Problématique de l'évaluation comme crise du sens	9
<i>Une mise en doute du jugement évaluatif des enseignants</i>	9
<i>Une évaluation contestée de l'« efficacité » des écoles</i>	10
<i>Une critique des évaluations dans l'enseignement supérieur</i>	11
Diversité des conceptualisations de l'évaluation en éducation : les enjeux de la modélisation	12
<i>Les paradigmes de l'évaluation</i>	13
<i>Les modèles d'évaluation</i>	14
<i>Les méthodes d'évaluation ou modèles d'action</i>	15
<i>Enjeux d'une réflexion critique sur la modélisation pour l'évaluation</i>	15
Organisation de l'ouvrage	16
<i>Les activités d'évaluation : quels savoirs produits ?</i>	16
<i>Les activités d'évaluation : quelles modélisations mises en œuvre ?</i>	18
<i>D'une épistémologie absente à une épistémologie en débat</i>	20
Références bibliographiques	21
LES ACTIVITÉS D'ÉVALUATION : QUELS SAVOIRS PRODUITS ?	25
La production de savoirs négociés dans deux recherches collaboratives sur l'évaluation formative	27
<i>Joëlle Morrissette*, Lucie Mottier Lopez** & Walther Tessaro**</i>	
<i>*Université de Montréal</i>	
<i>**Université de Genève</i>	
Une recherche identifiant le savoir de terrain en relation avec la mise en œuvre de l'évaluation formative	29
<i>Un modèle qui organise la pratique :</i>	
<i>l'exemple d'une routine d'évaluation formative</i>	30
<i>Une théorisation du savoir enseignant en matière d'évaluation formative</i>	31

Une représentation schématique comme levier à la coconstruction de l'objet	32
Une analyse permettant de distinguer les « manières de faire » partagées, admises et contestées	32
Une théorisation du territoire symbolique de l'évaluation formative autour de trois zones de savoir	33
Une recherche collaborative étudiant les régulations interactives entre l'enseignant et ses élèves	34
<i>La coélaboration d'un questionnement de recherche sur la régulation interactive</i>	35
<i>Des analyses collectives des extraits d'interactions choisis par les enseignants</i>	36
<i>Quelques résultats produits sur les guidages formatifs des enseignants</i>	37
<i>Construction d'un modèle à partir de la recherche collaborative</i>	38
La recherche collaborative : quels éclairages sur l'évaluation formative ?	40
Références bibliographiques	42
Entre savoirs scientifiques et usages normatifs : les enquêtes internationales d'évaluation des acquis des élèves	45
<i>Dominique Lafontaine & Christian Monseur Université de Liège</i>	
Introduction	45
Les théories au fondement des enquêtes internationales	46
<i>Des théories relatives au domaine évalué : le cadre de référence</i>	47
<i>Des théories relatives à la mesure</i>	48
Les modèles de la réponse à l'item : fondements et implications	50
Les paramètres méthodologiques influents	51
<i>Définition de la population</i>	51
<i>Choix des textes et difficulté des tâches</i>	52
Les savoirs produits par les enquêtes internationales	54
<i>Un éventail de savoirs, de formes variées, pour des usages multiples</i>	54
<i>Des savoirs pour la recherche</i>	55
<i>Une double polarité : entre constat et norme</i>	55
<i>Des savoirs pour quels usages ?</i>	56
En guise de conclusion : la délicate posture du chercheur en évaluation	57
Références bibliographiques	59
Les évaluateurs, des épistémologues en actes ?	63
<i>Claire Tourmen & Patrick Mayen AgroSup Dijon</i>	
Les démarches professionnelles d'évaluation, une tentative de rupture avec les jugements spontanés de l'expérience	64
<i>Énoncer des jugements, un acte quotidien</i>	64
<i>Énoncer des jugements, un acte contestable</i>	65
<i>Énoncer des jugements, un acte professionnel ?</i>	65
L'analyse du travail d'évaluateurs dans plusieurs contextes	66

Résultats : les évaluateurs utilisent l'expérience des phénomènes comme ressource	67
<i>La prégnance des premiers jugements</i>	67
<i>La mise à distance des premiers jugements</i>	68
<i>Le recours à l'expérience antérieure des phénomènes</i>	68
<i>Les connaissances issues de l'expérience :</i> <i>des référents dans l'évaluation</i>	70
L'expérience de l'expérience des phénomènes, ou des épistémologies en actes	71
<i>Une prévention des biais pour « bien évaluer »</i>	71
<i>Des épistémologies en actes ?</i>	73
Conclusion	74
Références bibliographiques	75
LES ACTIVITÉS D'ÉVALUATION : QUELLES MODÉLISATIONS MISES EN ŒUVRE ?	79
L'évaluation des élèves.	
Une modélisation interactionniste des pratiques professorales	81
<i>Pierre Merle</i>	
<i>IUFM de Bretagne et Université européenne de Bretagne</i>	
L'approche docimologique	82
Le modèle déterministe	84
Le modèle interactionniste	88
<i>Différents types d'arrangements évaluatifs</i>	88
<i>Arrangements et biais sociaux d'évaluation</i>	91
<i>Arrangements évaluatifs et performances scolaires</i>	92
Conclusion	93
Références bibliographiques	95
Élaboration et validation d'un questionnaire d'attitudes en éducation : considérations épistémologiques et métrologiques	97
<i>Pierre Valois, Claude Houssemand & Astrid de Leeuw</i>	
<i>Université du Luxembourg</i>	
Introduction	97
Construit à mesurer et format général de l'outil de mesure	101
Génération des items : étude qualitative	103
<i>Échantillon</i>	103
<i>Instruments de mesure et procédure de collecte des données</i>	103
<i>Résultats</i>	104
<i>Considérations épistémologiques</i>	104
Prétest du questionnaire : estimation de la fidélité de l'échelle d'attitude	105
<i>Échantillon</i>	105
<i>Instruments de mesure et procédure de collecte des données</i>	105
<i>Résultats</i>	106
<i>Considérations épistémologiques</i>	106
Étude quantitative : validité de l'échelle d'attitude	107

<i>Échantillon</i>	107
<i>Instruments de mesure et procédure de collecte des données</i>	108
<i>Résultats</i>	109
<i>Considérations épistémologiques</i>	111
Conclusion	111
Références bibliographiques	112
Autorégulation et évaluation-soutien d'apprentissage	115
<i>Dany Laveault</i>	
<i>Université d'Ottawa</i>	
Objectifs de la contribution	115
Rôle de l'autorégulation	116
Modèle empirique d'autorégulation et validité théorique	118
Rôle de la régulation des émotions dans les stratégies d'EsA	124
Conclusion	126
Références bibliographiques	128
Les modes de pensée, organisateurs des modèles d'évaluation : pour la reconnaissance d'une épistémologie profane	131
<i>Michel Vial</i>	
<i>Université de Provence</i>	
Introduction	131
Le cadre théorique	131
Méthodologie de la recherche conduite	133
<i>Moment considéré par la recherche</i>	136
<i>Processus de modélisation et choix des modèles d'évaluation</i>	137
Les résultats	138
<i>Le cas d'Albert</i>	138
<i>Le cas de Joëlle</i>	138
<i>Le cas de Roberte</i>	139
<i>Le modèle crée l'objet</i>	140
Interprétation des résultats : pour la reconnaissance d'une épistémologie profane	141
Conclusion	143
Références bibliographiques	144
Modélisation de la participation des acteurs dans le processus d'évaluation de la formation	147
<i>Pedro Rodrigues* & Eusébio André Machado**</i>	
<i>*Université de Lisbonne</i>	
<i>**Université Portucalense</i>	
Un concept modélisateur : le processus de participation intégré à l'évaluation	147
Premier trajet de modélisation : dispositif informel d'évaluation	149
<i>La modélisation de départ : les modèles d'analyse qui encadraient la recherche</i>	149
<i>Modélisation de l'évaluation participative de la formation par les enseignants formés</i>	151

Deuxième trajet de modélisation : dispositif formel d'évaluation participative	154
<i>La modélisation comme point de départ : un référentiel pour la recherche</i>	154
<i>Modélisation de la participation des enseignants formés à l'évaluation de la formation</i>	156
Note finale : la recherche comme un processus de modélisation	158
Références bibliographiques	159
D'UNE ÉPISTÉMOLOGIE ABSENTE À UNE ÉPISTÉMOLOGIE EN DÉBAT	163
Enjeux épistémologiques, et donc aussi politiques, de la méta-évaluation	165
<i>Lise Demailly Université de Lille 1</i>	
Les objectifs cognitifs des méta-évaluations	166
La méthodologie des méta-évaluations	169
<i>Soutenir une distance critique</i>	169
<i>Saisir le dispositif d'évaluation comme une pratique sociale complexe</i>	170
<i>Prendre le dispositif d'évaluation comme un processus temporel</i>	171
<i>Repérer les problèmes de positionnement de la méta-évaluation</i>	171
Quelques résultats globaux des méta-évaluations	173
<i>Critique épistémologique de dispositifs particuliers</i>	173
<i>Efficacité des évaluations</i>	175
La pluralité des modes d'approche	175
Le caractère participatif de l'évaluation	176
<i>La nature de l'évaluation</i>	177
Conclusion : scientificité des méta-évaluations ?	178
Références bibliographiques	180
Les exigences inconciliables des activités évaluatives et leurs paradigmes de référence	181
<i>Linda Allal Université de Genève</i>	
Remarques préliminaires	182
Points de rencontre entre les champs de l'évaluation	183
Évaluation des dispositifs de formation	184
<i>Caractéristiques de l'évaluation des dispositifs de formation</i>	184
<i>Paradigmes de référence pour l'évaluation des dispositifs de formation</i>	185
Évaluation des apprentissages en situation de formation	187
<i>Caractéristiques de l'évaluation des apprentissages en formation</i>	188
<i>Paradigmes de référence pour l'évaluation des apprentissages en formation</i>	188
En conclusion	191
Références bibliographiques	193

À la recherche de paradigmes unificateurs ou de paradigmes intégrateurs ?	195
<i>Jean-Marie De Ketele</i>	
<i>Université catholique de Louvain</i>	
Quatre paradigmes princeps et quatre épistémologies	196
Les fondements conceptuels et idéologiques des quatre paradigmes	200
<i>L'évaluation intuitive pragmatique</i>	200
<i>L'évaluation mesure</i>	201
<i>L'évaluation de la maîtrise des apprentissages</i>	202
<i>L'évaluation pour les apprentissages</i> ou <i>l'évaluation-soutien d'apprentissage</i>	202
Vers un paradigme unificateur et intégrateur ?	204
Références bibliographiques	208
Configurations pour une épistémologie de l'évaluation	211
<i>Gérard Figari</i>	
<i>Université de Grenoble</i>	
Évaluation et pratiques scientifiques	212
<i>En préalable, des fondements théoriques sous-jacents</i> <i>aux pratiques d'évaluation</i>	212
<i>Des prises de position épistémologiques</i>	213
<i>La question de la dimension scientifique des pratiques d'évaluation</i>	214
Des critiques de l'assujettissement de l'évaluation au positivisme systématique	214
Une distinction « évidente » entre pratique évaluative et recherche sur l'évaluation ?	214
<i>La question de la relation de l'évaluation au savoir :</i> <i>quel(s) savoir(s) ?</i>	216
Savoir ou connaissance ?	216
Évaluation, production de savoirs, construction de connaissances	216
La participation des acteurs de terrain à l'élaboration de savoirs	217
<i>Le « village des évaluateurs »</i>	218
Modèles et modélisations au service de l'évaluation	218
<i>Des opportunités de configurations plutôt qu'un éclatement</i> <i>de « modèles » ou de modélisations</i>	218
<i>Une configuration de niveaux de modèles d'évaluation</i> <i>reconnaissables ?</i>	220
Des paradigmes d'évaluation en guise d'épistémologie ?	
<i>L'invocation de « paradigmes d'évaluation »</i>	220
Le paradigme scientifique	221
Le paradigme comme vision du monde	221
Le paradigme comme ensemble de modèles d'évaluation	221
<i>Le recours aux « paradigmes » : un lien inéluctable</i> <i>entre recherche et évaluation ?</i>	222
Pour conclure	222
Références bibliographiques	224
Notices biographiques	225
Table des matières	233

Collection Raisons Éducatives

Concept éditorial

Raisons éducatives est une collection de la Section des sciences de l'éducation de l'Université de Genève publiée en partenariat avec les Éditions De Boeck Université.

Chaque édition traite un thème de façon approfondie, avec des regards multiples permettant la controverse scientifique, en regroupant des travaux effectués aussi bien dans la section que dans d'autres universités.

Raisons éducatives vise à...

- promouvoir un support scientifique de référence et de langue française, pouvant en principe couvrir l'ensemble des domaines de la recherche en éducation, et adressé en particulier à la communauté des chercheurs et des étudiants en sciences de l'éducation ;
- privilégier les questions vives des théorisations éducatives, en développant une approche restituant la diversité disciplinaire des sciences de l'éducation, ainsi que leur spécificité et leur inscription plus large dans le domaine des sciences humaines et sociales ;
- contribuer à donner forme, au fur et à mesure des publications et selon les cas :
 - à un repérage des objets et concepts émergents,
 - à un approfondissement des apories ou difficultés repérées comme majeures dans le champ,
 - à une ouverture de « chantiers » réflexifs nouveaux ;
- développer les liens scientifiques entre universités sur des thèmes significatifs au plan des questions et des hypothèses de recherche.

Comité de rédaction

Le travail principal du comité de rédaction est de définir le thème retenu pour chaque publication de la collection, de repérer et susciter les contributions souhaitées ainsi que de veiller au respect de l'orientation éditoriale.

Laurent FILLIETTAZ & Lucie MOTTIER LOPEZ (responsables du comité de rédaction), Abdeljalil AKKARI, Kristine BALSLEV, Jean-Michel BAUDOUIN, Fernando CARVAJAL SÁNCHEZ, Isabelle COLLET, Caroline DAYER, Joaquim DOLZ, Jean-Luc DORIER, Georges FELOUZIS, Gérard FIGARI, Siegfried HANHART, Francia LEUTENEGGER, Florence LIGOZAT, Valérie LUSSI, Jean-Paul PAYET, Christophe RONVEAUX, Bernard SCHNEUWLY, Sabine VANHULLE.

ADRESSE : RAISONS ÉDUCATIVES

Université de Genève – Faculté de psychologie et des sciences de l'éducation – Uni-Pignon, 40 Bd du Pont-d'Arve, CH-1211 Genève 4

Courrier électronique : publications-ssed-infos@unige.ch

Secrétariat : Leïla LOUCA

Comité scientifique

Le comité scientifique regroupe des chercheurs en éducation suisses et étrangers. La composition de ce comité est représentative des champs disciplinaires et des domaines de référence propres aux sciences de l'éducation. Chaque contribution de la série est soumise, pour expertise, à deux lecteurs du comité scientifique.

Linda ALLAL, Université de Genève (Suisse) – Jacques ARDOINO, Université Paris VIII Saint-Denis (France) – Réjean AUGER, Université du Québec à Montréal (Canada) – Daniel BAIN, Service de recherche en éducation (SRED), Genève (Suisse) – Elisabeth BAUTIER, Université Paris VIII Saint-Denis (France) – Véronique BEDIN, Université de Toulouse II (France) – Jean-Paul BERNIÉ, Université de Bordeaux II (France) – Pascal BRESSOUX, Université Pierre Mendès-France, Grenoble (France) – Bernadette CHARLIER, Université de Fribourg (Suisse) – Yves CHEVALLARD, Institut universitaire de formation des maîtres, Aix-Marseille (France) – Marcel CRAHAY, Université de Genève (Suisse) – Lucien CRIBLEZ, Université de Zurich (Suisse) – Bertrand DAUNAY, Université de Lille III (France) – Jean-Marie DE KETELE, Université catholique de Louvain (Belgique) – Christian DEPOVER, Université de Mons (Belgique) – Jean-Louis DUFAYS, Université catholique de Louvain (Belgique) – Vincent DUPRIEZ, Université catholique de Louvain (Belgique) – Michel FABRE, Université de Nantes (France) – Georges FELOUZIS, Université de Genève (Suisse) – Laurent GAJO, Université de Genève (Suisse) – Hubert GASCON, Université du Québec à Rimouski (Canada) – Monica GATHER THURLER, Université de Genève (Suisse) – Itziar IDIAZABAL, Université du Pays Basque (Espagne) – Anne JORRO, Université de Toulouse II (France) – Samuel JOSHUA, Université de Provence, Marseille (France) – Mokhtar KADDOURI, Université de Lille I (France) – Licinio LIMA, Université du Minho (Portugal) – Philippe LOSEGO, Haute École Pédagogique, Vaud (Suisse) – Régis MALET, Université de Bordeaux IV (France) – Claire MARGOLINAS, Université de Clermont-Ferrand (France) – Olivier MAULINI, Université de Genève (Suisse) – Patrick Mayen, AgroSup, Dijon (France) – Nathalie MONS, Universités de Grenoble et de Dijon (France) – Christian MONSEUR, Université de Liège (Belgique) – Danielle Moore, Université Simon Fraser, Vancouver (Canada) – Christiane MORO, Université de Lausanne (Suisse) – Nicole MOSCONI, Université Paris X Nanterre (France) – Tania OGAY, Université de Fribourg (Suisse) – Léopold PAQUAY, Université catholique de Louvain (Belgique) – Pierre PASTRÉ, Conservatoire national des arts et métiers (France) – Greta PELGRIMS, Université de Genève (Suisse) – Danièle PÉRISSET, Haute Ecole Pédagogique, Valais (Suisse) – Anne-Nelly PERRET-CLERMONT, Université de Neuchâtel (Suisse) – André PETITAT, Université de Lausanne (Suisse) – Geneviève PETITPIERRE, Université de Fribourg (Suisse) – Patrick PICARD, Institut français d'éducation (France) – Xavier PONS, Université Paris-Est Créteil (France) – Marie-Christine PRESSE, Université de Lille I (France) – Patricia REMOUSSENARD, Université de Lille III (France) – Bernard REY, Université Libre de Bruxelles (Belgique) – Olivier REY, Institut français d'éducation (France) – André ROBERT, Université Lumière Lyon II (France) – Frédéric SAUJAT, Université de Provence, Marseille (France) – Gérard SENSEVY, Université de Bretagne occidentale (France) – Georges SOLAUX, Université de Bourgogne (France) – Maurice TARDIF, Université de Montréal (Canada) – Andrée TIBERGHIEU, Université Lumière Lyon II (France) – Catherine VAN NIEUWENHOVEN, Université catholique de Louvain (Belgique) – Gérard VERGNAUD, Centre national de la recherche scientifique (France) – Richard WITORSKI, Université de Rouen (France).