

POLITIQUES &
MANAGEMENT
PUBLIC

Politiques et management public

Vol 31/1 | 2014

L'évaluation des politiques publiques

Contributions des sciences de l'éducation à la compréhension de la pratique évaluative

Education sciences' contribution to the understanding of evaluation practice

Claire Tourmen



Electronic version

URL: <http://journals.openedition.org/pmp/6970>

ISSN: 2119-4831

Publisher

Institut de Management Public (IDPM)

Printed version

Date of publication: 1 January 2014

Number of pages: 69-85

ISBN: 978-2-7430-1995-2

ISSN: 0758-1726

Electronic reference

Claire Tourmen, « Contributions des sciences de l'éducation à la compréhension de la pratique évaluative », *Politiques et management public* [En ligne], Vol 31/1 | 2014, mis en ligne le 01 décembre 2016, consulté le 19 avril 2019. URL : <http://journals.openedition.org/pmp/6970>

Contributions des sciences de l'éducation à la compréhension de la pratique évaluative

04

► **Claire Tourmen**

*Maître de Conférences en Sciences de l'Éducation
AgroSup Dijon, 26 Bd du Dr PetitJean, BP87999, 21079 Dijon Cedex*

Résumé

Cet article vise à synthétiser certains apports des recherches en sciences de l'éducation au champ de l'évaluation de programmes et à la connaissance de l'évaluation. Nous nous centrons plus particulièrement sur des recherches qui abordent l'évaluation comme une pratique et explorent l'activité des évaluateurs, tentant de décrire et de comprendre son déroulement. Nous montrons que plusieurs modélisations de l'activité évaluative ont été proposées en éducation et que les études sur l'activité évaluative en situation ont permis de dégager 4 grands constats : 1) les jugements sont construits de manière progressive et dynamique, 2) une partie du processus demeure implicite, 3) évaluer est une activité complexe faite d'adaptations et d'arbitrages, 4) évaluer peut s'exercer de manière plus ou moins experte. Cet article vise finalement à illustrer une partie des apports possibles des recherches en éducation à une théorisation plus globale de l'évaluation.

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

Mots clés : sciences de l'éducation, évaluation de programme, activité, jugement.

Abstract

Education sciences' contribution to the understanding of evaluation practice. In this paper, we will draw a synthesis of what research projects in the field of education can bring to the field of program evaluation and to the knowledge of evaluation itself. We will focus on research projects in which evaluation is seen as a practice and in which evaluators' activities in real situations are explored, in order to be described and understood, to highlight their specificities and their main logic. We will show that different models of evaluative activity have been proposed in education and that the studies

*Auteur correspondant : Claire.tourmen@educagri.fr

doi :10.3166/pmp.31.69-85 © 2014 IDMP/Lavoisier SAS. Tous droits réservés

on evaluative activities have highlighted 4 features of them: 1) judgments are build in a progressive and dynamic way, 2) part of the process remains implicit, 3) evaluating is a complex activity made of lots of adaptations and compromises, 4) evaluating can be done in a more or less expert way. This paper finally aims at illustrating possible lessons learned from research on evaluation in education, and their possible contribution to a general theory of evaluation.

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

Keywords: education sciences, Program evaluation, activity, judgment.

Introduction

La préoccupation pour l'évaluation est chose ancienne dans le champ éducatif et elle a précédé les préoccupations autour de l'évaluation de programmes, historiquement plus récentes. L'évaluation de programmes ou de politiques publiques¹ s'est développée à partir des années 1960 et donne lieu aujourd'hui à de nombreuses études dans différents champs (sciences politiques, sociologie, économie, sciences de gestion...). Les pratiques d'évaluation des apprentissages sont plus anciennes et les premières études sur l'évaluation ont commencé dès les années 1950 en éducation et en psychologie, comme le rappellent Guba et Lincoln (1989). L'évaluation des apprentissages, des compétences, des dispositifs de formation et des politiques éducatives reste un sujet de préoccupation pour les chercheurs en sciences de l'éducation ainsi que pour les praticiens du domaine éducatif (enseignants, inspecteurs, responsables de formation, consultants...), témoin le succès annuel du colloque international de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (Admee²) et sa revue scientifique, *Mesure et Évaluation en Éducation*.

Notre propos sera de tenter de montrer certains apports des sciences de l'éducation à la compréhension de l'évaluation de programmes, et, au-delà, de toute pratique d'évaluation. Un des apports majeurs que nous développerons ici part de l'idée qu'évaluer est avant tout une activité pratique avant d'être un discours, et que cette pratique peut être observée, analysée, modélisée voire théorisée. Nous verrons tout d'abord que plusieurs auteurs en éducation ont tenté, depuis les années 1950, de modéliser les opérations invariantes de la pratique évaluative dans le contexte éducatif, voire de toute pratique évaluative, quels que soient ses objets. Des recherches plus récentes en éducation ont permis de compléter et de discuter ces modèles par des observations de l'activité d'évaluateurs *in situ*. Nous détaillerons ensuite, sur la base de données issues de plusieurs recherches, ce que de telles modélisations ont pu apporter à la compréhension de l'activité évaluative, qu'il s'agisse d'enseignants ou d'évaluateurs de programmes et politiques publiques. Nous concluons sur les apports possibles des sciences de l'éducation à une théorisation plus globale de l'évaluation, en complémentarité avec d'autres disciplines.

¹ Le premier terme est celui qui est largement utilisé en Amérique du Nord (y compris au Québec) tandis que le second est davantage utilisé en France. Nous ne rentrerons pas ici dans les distinctions qui sont données.

² www.admee.org.

1. L'évaluation, un objet de préoccupation ancien et fécond en éducation

1.1. Ancienneté du propos sur l'évaluation en éducation

Les études sur l'« évaluation » ont émergé dans la première moitié du XX^e siècle, avant même la généralisation de pratiques d'évaluation appliquées aux programmes et politiques publiques à partir des années 1960 aux États-Unis (voir à ce sujet Shadish, Cook et Leviton, 1991, ou Monnier, 1992)³. Les premières études mentionnant l'évaluation portaient sur la mesure des apprentissages, de l'intelligence et des facultés psychologiques, notamment par la mise au point de tests psychométriques, comme le rappellent Guba et Lincoln (1989). Cela correspond à ce qu'ils appellent « la première génération » et la « deuxième génération de l'évaluation », au milieu du XX^e siècle. Les champs de l'éducation et de la psychologie ont ainsi une certaine antériorité sur le discours sur l'évaluation, discours qui n'a cessé de se développer depuis dans un nombre de plus en plus grand de disciplines au fur et à mesure de l'élargissement des champs couverts par les pratiques évaluatives. Le développement parallèle du champ de l'évaluation en éducation et du champ de l'évaluation de programmes et politiques publiques⁴ laisse parfois l'impression d'un manque de dialogue entre eux. Malgré tout, quelques chercheurs nord-américains ont servi de passerelles entre ces deux champs (Scriven, 1967, Stufflebeam, 1980⁵) et deux tiers des formations universitaires à l'évaluation de programme restent réalisées, à l'échelle des États-Unis, dans les départements d'éducation ou de psychologie (Lavelle et Donaldson, 2010). Lavelle et Donaldson l'expliquent notamment par le fait que des psychologues et des pédagogues se sont impliqués, depuis les années 1960, dans le développement des pratiques et des méthodes d'évaluation de programme. Ils rappellent aussi que parmi les premiers programmes appelés à être évalués de manière systématique aux États-Unis, au début des années 1960, figuraient des programmes éducatifs (voir aussi Shadish, Cook et Leviton, 1991), et que les programmes éducatifs restent beaucoup évalués à ce jour.

1.2. L'étude de l'évaluation en éducation

Les premières recherches sur l'évaluation en éducation portaient sur la mesure des apprentissages et les conditions de sa validité, l'évaluation étant alors définie comme un processus de mesure (Tyler, 1950, Bonboir, 1972). Puis, à partir des années 1970, de nouvelles définitions de l'évaluation ont émergé en éducation. L'évaluation a été définie comme une production de jugements finalisés par une utilité sociale (Stufflebeam, 1980). Ardoino observait ainsi, en 2001, en faisant le bilan de plusieurs décennies de recherches sur l'évaluation dans le domaine éducatif, « l'élargissement du champ de l'évaluation » et sa « complexification »

³ Même si des premières études qui peuvent être apparentées à des évaluations de programme ont eu lieu dans la première moitié du XX^e siècle, comme le rappelle Monnier (1992), elles sont restées isolées.

⁴ Par exemple, en France, l'association Admee et la Société Française de l'Évaluation n'ont entamé que très récemment (2011) un rapprochement par le biais d'un séminaire conjoint (1^{er} décembre 2011, « Évaluation en éducation. Qui évalue et sur quels fondements ? ». http://www.sfe-asso.fr/sfe-evaluation.php?menu_id=852&PHPSESSID=131d55033b876506d5292478da881a02)

⁵ Ces deux auteurs, issus du champ de l'éducation, sont devenus célèbres dans celui de l'évaluation de programmes.

depuis les premières études. « Au lieu de se contenter longtemps des questions « comment ? », voire « pourquoi ? », on commence à s'interroger sur les « pour quoi » et « pour qui » (2001, p. 227). De plus en plus d'études ont mis l'accent sur les enjeux des résultats de la mesure (Merle, 2007, Demailly, 2001, Dauvisis, 2006) et pas uniquement sur le processus technique lié à l'observation des objets. L'accent a été mis sur les effets des évaluations, notamment en termes de métacognition et d'impact sur les apprentissages (Allal, 2001). Des pratiques d'auto-évaluation et de co-évaluation ont ainsi été testées (Paquay, Darras et Saussez, 2001). Toutefois, comme le notait Figari en 2006, cette distinction entre mesure et jugement fait toujours débat, alimentant les polémiques autour des conceptions de l'évaluation.

À cette multiplication des définitions s'est ajoutée l'attribution de différentes fonctions à l'évaluation, une des plus célèbres catégorisations étant celle de Scriven (1967) qui distingue des fonctions « sommatives » (certification et reddition de comptes) des fonctions « formatives » (régulation et apprentissage). D'autres catégorisations des fonctions de l'évaluation ont été produites (Cardinet, 1986, De Ketele, 2006), certaines insistant sur l'opposition entre fonction de « contrôle » et fonction de « régulation » de l'évaluation (Ardoino et Berger, 1989, Vial, 2001), voire en proposant de classer différentes « postures » d'évaluateurs (Jorro, 2006). Les différentes approches de l'évaluation en éducation ont enfin été classées en « paradigmes » concurrents, terme repris à Kuhn (1972) pour désigner différents systèmes de « visions du monde » (Ardoino, 2001). Cardinet (1986) et Rodrigues (2001) ont ainsi respectivement différencié trois paradigmes de l'évaluation en éducation : le paradigme « externe » ou « objectiviste »⁶, le paradigme « interne » ou « subjectiviste »⁷, enfin le paradigme « négocié » ou « dialectique »⁸, ces différents paradigmes renvoyant à différents modes d'association des évalués soutenus par différentes fonctions attribuées à l'exercice (Tourmen, 2007, p. 78).

Les chercheurs en éducation ont donc eu à cœur, depuis longtemps, de définir l'évaluation, d'en classer les fonctions et d'en dériver des considérations méthodologiques et épistémologiques (Mottier-Lopez et Figari, 2012) permettant de prescrire certaines formes d'évaluation, puis de classer ces formes d'évaluation en « paradigmes ». Il est intéressant de noter que l'on observe un mouvement similaire dans la littérature internationale en évaluation de programmes ou politiques publiques (Tourmen, 2007), bien que cette dernière renvoie à d'autres champs disciplinaires (sociologie, science politique, économie, pour ne citer qu'eux). Différentes définitions de l'évaluation se sont également succédé (Guba et Lincoln, 1989), différentes fonctions ont été attribuées au processus (CSE, 1996, Baslé, 2008), différents paradigmes ont été distingués (Shadish, Cook et Leviton, 1991, Monnier, 1992, Perret, 2001), menant à ce que Patton (1986) a appelé « la guerre des paradigmes » autour des méthodes et des finalités préconisées pour l'exercice évaluatif.

1.3. Modèles de l'activité évaluative

Nous allons à présent centrer notre propos sur ce qui nous semble constituer l'un des apports importants, si ce n'est majeurs, de l'étude de l'évaluation en éducation, à savoir l'étude de l'activité évaluative. Depuis les travaux fondateurs de Tyler (1950) et des docu-

⁶ Mesure à fin de vérification, évaluateurs externes aux évalués.

⁷ Finalité d'apprentissage, auto-évaluation par les évalués.

⁸ Finalité de régulation, participation négociée de tous les acteurs (évaluateurs, évalués).

mologues (étude de la notation, voir Bonboir, 1972, pour une synthèse), plusieurs auteurs ont en effet tenté de modéliser le déroulement de l'activité évaluative dans le contexte éducatif, appliquant leurs modèles à l'évaluation des apprentissages (Cardinet, 1986, Bonniol, 1981, Bonniol et Vial, 1997), à l'évaluation des dispositifs éducatifs (Barbier, 1985, Figari, 1994, 2014) et enfin à celle des programmes éducatifs (Scriven, 1981, Stufflebeam, 1980). Certains de ces auteurs ont même proposé d'étendre la validité de leurs modèles à toute activité évaluative, quel que soit son domaine (pas uniquement l'éducation) ou l'échelle de l'objet évalué (micro, comme les apprentissages d'un sujet, méso, comme un dispositif, un projet ou un programme délimité, ou macro, comme une politique publique).

Les auteurs de ces études ont défini la pratique évaluative comme une activité ou un processus qui serait plus qu'une simple procédure déroulant des méthodes et des techniques. Figari a constaté un tel mouvement dans l'introduction d'un ouvrage synthétisant les recherches en évaluation en éducation (Figari et Achouche, 2001, p. 10) : « une caractéristique a paru assez rapidement commune à la plupart des textes : celle consistant à parler de l'évaluation en terme de processus en ne présentant les procédures que comme des méthodes utilisées par les acteurs de l'évaluation. L'utilisation du terme « activité évaluative », à la place de celui d'évaluation, s'est alors imposée, voulant indiquer que la tendance majeure des évolutions constatées consiste, à travers les pratiques, les problématiques et les méthodes, à s'intéresser particulièrement à ce qui se passe dans l'acte d'évaluer. »

Nous allons résumer les principaux traits de ces modèles qui assimilent l'évaluation à une activité de comparaison réalisée en plusieurs étapes :

- **Tyler (1950)** définissait l'évaluation comme une mesure qui prend la forme d'une comparaison objectifs-résultats : « Le processus d'évaluation consiste essentiellement à déterminer dans quelle mesure les objectifs d'éducation sont en voie d'être atteints par le programme d'étude et de cours. Cependant, comme les objectifs d'évaluation visent essentiellement à changer les êtres humains, c'est-à-dire que le but est de produire dans le comportement des étudiants certains changements consistant à déterminer dans quelle mesure ces changements de comportement sont en train de se produire » (1950, p. 69). Cette définition a inspiré de nombreux travaux en éducation qui ont depuis porté sur la mesure des apprentissages et l'étude de ses conditions de fiabilité/validité⁹.
- **Scriven (1981)** a élargi le propos de Tyler et proposé un modèle de la « logique évaluative », toute activité d'évaluation pouvant se résumer à quatre étapes fondamentales : 1) établir des critères, 2) construire des seuils de réussite, 3) rassembler des données, 4) intégrer les résultats dans le jugement final. Il s'agit d'un des modèles les plus cités dans la communauté anglophone des évaluateurs. Scriven y définit l'évaluation comme la production de jugements au sujet du « mérite » et de la « valeur » des objets.
- **Stufflebeam (1980)** a également proposé un modèle valable pour toute activité évaluative. Il a commencé par préciser que, dans tout jugement évaluatif, on retrouve la même structure, même si celle-ci reste implicite comme dans le jugement rapide de la vie quotidienne : « Parce qu'il est implicite, ce processus garde secrets à la fois

⁹ Que ce soit la docimologie (étude de la notation et de ses biais, voir Bonboir, 1972), l'éduométrie (Cardinet, 1986) ou étude « scientifique » de la mesure des apprentissages, et plus récemment les travaux portant sur les indicateurs de mesure de « l'école efficace » (Demeuse, 2006).

les données considérées et les critères ou standards qui ont servi à les apprécier » (1980, p. 17). Il a ensuite proposé une définition – souvent reprise - de l'évaluation « formelle » : « L'évaluation est le processus par lequel on délimite, obtient et fournit des informations utiles permettant de juger des décisions possibles » (1980, p. XXVI). Il a ensuite identifié trois étapes toujours présentes dans un processus d'évaluation : « Délimiter » l'information nécessaire, « Obtenir » l'information et « Fournir », ou communiquer, l'information (1980, p. 275).

- **Cardinet (1986)** a lui aussi défini l'évaluation comme un processus de mesure réalisé en trois étapes : « Recueillir l'information », « Interpréter l'information » et « Utiliser l'information ». Il observe que « ces trois phases sont étroitement liées : on n'observe que pour pouvoir interpréter et immédiatement agir » (1986, p. 38).
- **Bonniol (1981)**, dans sa thèse sur l'évaluation des apprentissages (1981, rapportée dans Bonniol et Vial, 1997, pp. 63-65), a également tenté de définir un modèle des actions structurantes de l'activité évaluative. Selon lui, « toute évaluation porte sur un comportement » défini par des objectifs. L'évaluateur les reconstruit en définissant des critères d'évaluation, « il s'agit des dimensions, des caractéristiques qui seront privilégiées lors de l'évaluation, et qui définissent le produit norme dans l'ensemble des produits possibles ». L'évaluateur détermine ensuite le « produit attendu » et y applique une échelle d'étalonnage, ce qui donne le « modèle de référence ». « L'évaluation est alors définie comme une comparaison entre un produit réel et un modèle de référence, comme un jugement comparatif ». Cette démarche est itérative au sens où le « modèle de référence » peut être ajusté aux vues des « produits réels ». Bonniol ajoute que ce « modèle simple » de l'évaluation « permet a priori d'analyser n'importe quel dispositif d'évaluation » (1981). « Réduisant sans doute la réalité qu'il représente de manière globale et simplifiée, il distingue et met en relief des rouages trop souvent confondus ou au contraire envisagés dans une indépendance fallacieuse : objectifs d'évaluation, critères, normes, barèmes... » Ce modèle a pour ambition de « décrire la structure de l'évaluation » (1981).
- **De Ketele (1985)** a repris et élargi la définition initiale de Stufflebeam (1980). Selon la nouvelle définition qu'il propose, beaucoup citée en éducation, évaluer consiste à « recueillir un ensemble d'informations suffisamment pertinentes, valides et fiables et examiner le degré d'adéquation entre cet ensemble (d'informations valides, pertinentes et fiables) et un ensemble de critères adéquats aux objectifs fixés au départ ou ajustés en cours de route, en vue de prendre une décision ».
- **Barbier (1985)** a lui aussi tenté de modéliser l'évaluation, cette fois-ci appliquée aux actions de formation. Comme Stufflebeam (1980), il est parti de l'opposition entre l'évaluation « instituée »¹⁰ et l'évaluation « spontanée », où « le jugement de valeur ne s'explique qu'à travers son énoncé, sa formulation » (Barbier, 1985, p. 34). Selon lui, l'activité d'évaluation « instituée » consiste premièrement à produire des « référents » (1985, p. 83), ce qu'il appelle « la tâche d'élucidation des référents » et de « spécification des critères » (1985, p. 77). Cette tâche précède la « production des données » ou « production des référés », elle-même précédant leur comparaison.

¹⁰ Ou « acte délibéré et socialement organisé aboutissant à la production d'un jugement de valeur » (1985, pp. 33-34).

- **Figari (1994, 2014)** a enfin lui aussi proposé une modélisation des actes impliqués dans toute activité d'évaluation à travers son modèle de la « référentialisation », appliqué à l'évaluation de dispositifs mais applicable à d'autres objets d'évaluation (Tourmen et Figari, 2006). Toute évaluation commence, selon lui, par la construction/explicitation de « référents » (éléments permettant d'attribuer une valeur aux « référés »), ce qu'il a appelé une phase de « référentialisation ». Cette phase consiste à construire un référentiel d'évaluation qui explicite les référents et les référés recherchés, elle est suivie d'une phase de récolte de données (« référés ») sur l'objet évalué puis d'une comparaison des « référés » aux « référents », ce qui constitue l'acte de jugement lui-même, permettant d'attribuer une valeur à l'objet évalué. Figari a insisté sur l'importance du « système de références », au pluriel, soulignant qu'« *il n'est pas envisagé d'évaluation sans recours implicite ou explicite à un système de références* » (1994, p. 31). Il a ajouté (2006 b) que les « systèmes de références » utilisés sont parfois explicites (comme c'est le cas lors du recours aux référentiels), parfois implicites, même s'ils sont toujours agissants.

Ces modélisations présentent plusieurs points communs :

- L'évaluation est décrite comme une activité séquencée organisée en étapes identiques ;
- Évaluer est défini comme une activité de comparaison entre ce qui est appelé des « référés » (données sur l'objet évalué) et des « référents » (normes ou informations servant de standards de comparaison) qui permettent de leur donner un sens, une valeur, à savoir de porter un jugement ;
- La comparaison impliquée dans le jugement final est précédée et rendue possible par 1) une étape préalable d'explicitation ou de construction des « référents », définis par des critères et rendus mesurables par des indicateurs (stabilisés par exemple dans un référentiel), puis par 2) une autre étape préalable de recueil des données sur l'objet évalué (ou « référés ») ;
- La démarche globale de production du jugement se base sur des actes explicités, montrés, maîtrisés, en opposition aux évaluations quotidiennes, spontanées et informelles, où ces actes restent implicites et soumis à de nombreux biais potentiels.

1.4. Études de l'activité évaluative en conditions réelles

Si ces modèles présentent un intérêt certain pour représenter l'activité évaluative, ils courent le risque de constituer des visions tronquées ou idéalisées de l'évaluation telle qu'elle est pratiquée dans la « vraie vie » (Bamberger, Rugh et Mabry, 2006). En effet, on ne sait pas toujours sur quelle base ces modèles ont été construits (des observations ? une pratique ? des intuitions ?). De plus, quantité d'écrits en évaluation étant davantage de nature prescriptive que descriptive, le doute demeure. C'est ce doute et l'envie de tester, d'affiner ces modèles qui ont récemment poussé plusieurs chercheurs à se doter d'observations d'évaluateurs au travail, en conditions réelles, pour aller plus loin dans la connaissance de l'activité évaluative. Ces chercheurs retrouvent en un sens les préoccupations initiales des docimologues qui basaient aussi leurs études, dans les années 1960 et 1970, sur l'analyse de l'activité évaluative (en laboratoire notamment, voir Bonboir, 1972). Ainsi, Merle a finement observé le déroulement des activités d'évaluation en classe et dans des jurys de baccalauréat (2007), Demailly a observé le déroulement d'activités d'évaluation dans des

établissements scolaires (2001), Hurteau a testé des modèles de l'activité évaluative en observant des rapports d'évaluation (Hurteau, Houle et Mongiat, 2009) puis des activités de production de jugement dans différents domaines professionnels (Hurteau, Houle et Guillemette, 2012), Allal et Mottier-Lopez ont interrogé des enseignants sur leurs pratiques d'évaluation (2009), et Tourmen a observé des évaluateurs de politiques publiques dans leur travail quotidien (2007, 2008, 2009b), ainsi que des enseignants professionnels en période d'évaluation (Tourmen, Deblay, Lipp et Boude, 2013).

Un mouvement similaire a été initié aux États-Unis, dans la littérature en évaluation de programmes, où de plus en plus de publications rapportent des études basées sur l'observation de l'activité réelle d'évaluateurs ou les encouragent (Fitzpatrick, 2004, Schwandt, 2005, 2008, Fitzpatrick, Christie et Mark, 2008, Donaldson et Lipsey, 2008, Tourmen, 2009, Kundin, 2010). Schwandt (2007), éditeur de la revue *American Journal of Evaluation*, appelle de ses vœux la poursuite de telles recherches, qui renouvellent la connaissance de l'évaluation et génèrent de nouvelles questions.

2. Évaluateurs au travail : acquis et questions de la recherche en éducation

Nous allons à présent synthétiser les résultats des études sur les évaluateurs au travail en illustrant notre propos avec des *verbatim* issus de données de deux d'entre elles. Ces études sont basées sur des observations d'évaluateurs *in situ* (Merle, 2007, Demailly, 2001, Tourmen, 2007, 2008, 2009b), sur des entretiens post activité avec des évaluateurs sur base de traces du travail (Allal et Mottier-Lopez, 2009, Hurteau, Houle et Guillemette, 2012) ou sur l'analyse de rapports d'évaluation (Hurteau, Houle et Mongiat, 2009). Elles mettent l'accent sur certains phénomènes permettant de contredire, confirmer et affiner, certaines hypothèses issues des modèles de l'activité évaluative exposés ci-dessus.

2.1. Quatre constats issus de la recherche en éducation

Nous avons résumé les principaux apports de ces recherches sur les évaluateurs au travail autour de quatre constats. Tout d'abord, **1) la construction des jugements est progressive et dynamique**. Évaluer peut certes se décrire comme une activité socialement organisée de production de jugements, mais il ne s'agit pas d'une activité séquencée, linéaire, ordonnée en stades indépendants. Au contraire, l'activité évaluative est itérative (ce qui confirme l'observation de Bonniol, 1981), faite d'allers et retours entre la construction des référés et celle des référents. Toutes les observations d'évaluateurs au travail montrent une construction progressive et dynamique des jugements. Cette construction passe par une « triangulation » progressive des informations sur base de plusieurs sources de données (Allal et Mottier Lopez, 2009), l'évaluation étant faite de multiples anticipations et d'allers et retours entre les hypothèses et les observations. La production des jugements finaux peut donc se décrire non pas comme une simple évaluation mais comme une multitude d'évaluations successives et provisoires s'ajustant progressivement au cours de la démarche (Tourmen, 2009a).

Ensuite, **2) une partie du processus demeure implicite**. Ces études confirment que les évaluateurs comparent bien des référents avec des référés à l'aide de critères et d'indicateurs, mais le processus demeure pour partie implicite (confirmant ici les observations de Figari,

1994). Ainsi, les critères et les indicateurs utilisés pour juger ne sont pas tous verbalisés ni totalement explicites (Hurteau, Houle et Mongiat, 2009, Tourmen *et al.*, 2013), une partie se découvre en cours d'activité et tous ne peuvent être prévus dans ses référentiels. Certains auteurs ont ainsi pu parler d'une « référentialisation implicite » (Tourmen et Droyer, 2011).

De plus, 3) évaluer est une activité complexe faite d'adaptations et d'arbitrages qui débordent les catégorisations de la littérature. Bien plus qu'une séquence d'étapes à dérouler et qu'un choix entre fonctions ou paradigmes nets et bien délimités, ces études montrent qu'il s'agit d'activités complexes et contextuelles (Hurteau, Houle et Guillemette, 2012). Évaluer nécessite de constantes adaptations des méthodes aux situations, de constants arbitrages entre des finalités multiples et parfois contradictoires (Merle, 2007), certains évaluateurs démontrant un usage souple et opportuniste des méthodes et des paradigmes selon les situations (Tourmen, 2007, Pons, 2010). Demailly (2001) affirme même que les catégorisations des fonctions de l'évaluation, au lieu de clarifier le débat, peuvent au contraire masquer certains enjeux. Elle conclut : « *Cela implique de ne pas accorder une attention excessive aux catégorisations méthodologiques produites en abondance dans toutes les approches technicistes de l'évaluation, lesquelles constituent l'essentiel de la littérature sur l'évaluation (évaluation interne/externe, sommative/formative...). (...) Elles (...) empêchent le plus souvent de saisir les relations de pouvoir et les enjeux effectifs qui se jouent dans un dispositif* » (2001, p. 22).

Enfin, 4) évaluer peut s'exercer de manière plus ou moins experte. Ces études montrent enfin qu'évaluer s'apprend (au risque du « bricolage », Dauvisis, 2006, Pons, 2010). Cette activité peut prendre différentes formes selon le profil et le degré d'expérience des évaluateurs (Hurteau, Houle et Guillemette, 2012, Tourmen, 2008). Selon la manière dont l'évaluation est conduite, elle peut avoir des effets inattendus voire pervers (Demailly, 2006). Même en déroulant correctement une méthode prescrite, un évaluateur peut en effet passer à côté de la complexité sociale de sa situation, renforcer les attitudes de rejet envers l'évaluation et corrompre les processus (d'apprentissage, de décision, de mise en œuvre d'une politique) qu'il pensait améliorer (Tourmen et Demailly, 2012).

Ces études soulignent finalement la complexité de l'activité évaluative en situation, son caractère dynamique et pour partie imprévisible. Même si elles ont été réalisées dans des contextes différents, elles révèlent ainsi d'autres dimensions invariantes de l'activité évaluative. Nous allons illustrer les quatre constats exposés ci-dessus à travers des données extraites de l'observation d'évaluateurs au travail dans deux recherches.

2.2. Étude de cas 1 : deux enseignants évaluant un élève

À travers un court extrait de l'activité de deux évaluateurs, nous allons illustrer les deux premiers phénomènes mis à jour, à savoir 1) la construction progressive des jugements et 2) le fait qu'une partie du processus demeure implicite.

Cet extrait montre une délibération à l'œuvre entre deux évaluateurs aboutissant à la production d'un jugement. Il est issu d'une recherche sur l'évaluation en situation professionnelle (Tourmen *et al.*, 2013) où nous avons réalisé l'observation (filmée) d'une matinée d'épreuves pratiques dans un lycée professionnel agricole, ainsi que des interviews pré et post activité avec les deux membres du jury et un échantillon d'élèves. Les deux évaluateurs sont une enseignante de zootechnie en lycée agricole et un professionnel de l'élevage. Tous deux ont organisé une épreuve pratique avec des élèves de seconde professionnelle dans

l'étable du lycée agricole. Les élèves ont dû réaliser des tâches comme « réaliser l'état de santé du troupeau » sous le regard des évaluateurs, qui délibèrent ensuite pour leur attribuer une note. Voici un extrait intégralement retranscrit de la délibération portant sur un élève :

Professionnel : *Ben sur le global moi j'aurais mis 15 quoi*

Enseignante : *Hum, moi j'ai trouvé qu'il s'est bien débrouillé, enfin je m'attendais à moins que ça et je suis contente de lui là, 15 ouais*

Professionnel : *Non ?*

Enseignante : *Oui, si si, oui parce que il, au niveau d'organiser et tout il a regardé et il a capté de lui-même*

Professionnel : *Bon après quand il l'a détachée du cornadi¹¹ pour lui mettre le licol¹² ça j'ai trouvé ça un peu...*

Enseignante : *Oui ça c'était un peu bizarre et pis si elle se barre en courant t'es content après* (rires)

Professionnel : *Voilà mais bon c'est vrai que c'est un peu dans le faux parce qu'ils ont des vaches un peu faciles...*

Enseignante : *Oui ils ont du mal à se dire...*

Professionnel : *Tu serais en charolais avec de la laitière ça serait pas du tout la même chose ».*

On constate tout d'abord que **les jugements évaluatifs sont bien le fruit d'une construction progressive et dynamique**. Un premier jugement – rapporté à une note – (« *sur le global moi j'aurais mis 15* ») a été énoncé avant d'être confirmé puis spécifié en étant rapporté à plusieurs critères et indicateurs (« *tout il a regardé* », « *il l'a détachée du cornadi* »). Il est intéressant de noter que certains critères n'ont été rapportés à aucun indicateur précis (« *bien débrouillé* »), l'accord entre évaluateurs semblant faire office de validation. Plusieurs critères ont ainsi été successivement examinés et mis en balance (« *bien débrouillé* », « *organiser* », « *capté de lui-même* », « *bizarre* »...), ils sont venus nourrir l'ajustement du jugement initial.

Sur plusieurs autres terrains, la présence de jugements précoces, provisoires et partiels a été observée, dès les étapes de conception d'une évaluation et dès les premiers moments de l'observation. Ces jugements peuvent acquérir – s'ils sont mis en doute – le statut d'hypothèses de travail dans la suite de l'activité (Tourmen, 2009a). Ces jugements sont, dans le cas d'évaluations collectives, énoncés puis mis en débat, argumentés, rapportés à des indicateurs et critères différents, avant d'être stabilisés. Ceci confirme les résultats de Alla et Mottier-Lopez (2009) qui ont observé un phénomène similaire à l'œuvre chez des enseignants du primaire, phénomène qu'elles ont nommé « triangulation » en référence à des études sur la construction de la connaissance. Hurteau, Houle et Guillemette (2012) ont également observé que « *le processus de production de l'information est constitué d'une série de boucles et de va-et-vient entre la collecte et leur traitement* » (2012, p. 90) et que, « *peu à peu les différentes versions du jugement se raffineront* » (p. 95).

On constate aussi, dans l'extrait exposé ci-dessus, qu'**une partie du processus de jugement reste implicite**. Les évaluateurs pondèrent leur jugement progressif au regard des différents référents qu'ils mobilisent, dont deux ont été énoncés ici : la connaissance des niveaux de difficulté

¹¹ Dispositif autobloquant qui permet aux animaux de rester à l'auge.

¹² Pièce de harnachement pour les animaux domestiques qui se place sur le cou.

du travail dans d'autres situations (« *Tu serais en charolais avec de la laitière ça serait pas du tout la même chose* ») ainsi que le niveau attendu d'un élève connu (« *je m'attendais à moins que ça* »). Ces référents n'ont pas été énoncés avant l'évaluation, ni discutés entre évaluateurs, mais ils ont été visiblement actifs. Ils sont finalement apparus en cours de délibération. Une partie des référents (qu'entendent-ils par « *bien débrouillé* »?) est en revanche restée implicite.

Ces résultats vont dans le sens des travaux de Merle (2007) sur la notation chez les enseignants, et, avant lui, des travaux en docimologie (Bonboir, 1972) et en psychologie sociale (Bressoux et Pansu, 2003) sur les biais de notation. Il apparaît que, souvent, les évaluateurs recourent à plusieurs systèmes de référence, certains provenant de leur expérience (Mayen, Mayeux et Savoyant, 2006, sur les jurys de VAE), et qu'une partie de ces systèmes de référence reste implicite, bien qu'agissante. Demailly (2001) a quant à elle observé la relativité des systèmes de référence qui guident les évaluateurs en établissements scolaires : « *Les évaluations sont foncièrement des opérations de compromis entre acteurs attachés à des référentiels de l'action publique différents.* » (p. 23), comme un référentiel d'obligation de moyens, un modèle de régulation néo-libéral... Elle observe la coexistence de ces différentes visions du monde dans les pratiques éducatives. De plus, tous les jugements d'un même évaluateur ou d'un même jury ne se basent pas systématiquement sur les mêmes critères et ni sur les mêmes indicateurs, d'où des questions d'équité plus spécifiques aux évaluations en éducation, comme le souligne Belair (1990). L'une des vertus des jurys collectifs est de forcer les évaluateurs à ajuster leurs références et les critères pris en compte pour limiter les biais. Selon Merle (2007), « *le sens des délibérations est de surmonter les évaluations divergentes et d'élaborer un modus vivendi qui satisfait les différentes parties prenantes* » (p. 164). Selon Hurteau, Houle et Guillemette (2012), un jugement crédible « s'appuie sur une question d'évaluation, des critères et des seuils de réussite » (p. 5) qu'il est préférable d'explicitier, du moins de discuter.

Un moment de délibération comme celui-ci – dans le cas de la construction collective d'un jugement entre plusieurs évaluateurs – est donc un moment intéressant à observer car on peut suivre pas à pas la formation du jugement final à partir de jugements précoces énoncés puis discutés. On peut suivre aussi l'émergence progressive de différents critères, leur degré de précision et les indicateurs avec lesquels ils sont mis en relation, ainsi que leur pondération. On observe aussi les règles de parole et le poids de chaque évaluateur dans le processus, l'énoncé et l'éventuelle mise à distance de jugements initiaux ou de pré-jugements... Une telle analyse permet de mieux comprendre ce qui se passe, d'étudier la manière dont critères et indicateurs sont convoqués, utilisés et lesquels le sont (ou non), autrement dit, d'étudier la « référentialisation » implicite à l'œuvre (Figari, 1994, Tourmen et Figari, 2006, Tourmen et Droyer, 2011). On peut par exemple :

- mettre à jour la différence entre critères prescrits – officiels, comme dans le référentiel de certification ici – et critères effectivement pris en compte pour juger – qui reposent notamment sur les conceptions propres aux évaluateurs –,
- discuter le poids des attentes et les biais possibles dans la construction des jugements (« *je m'attendais à moins que ça* »), dans la lignée des premiers travaux de docimologie (Bonboir, 1972),
- mettre à jour le référentiel réel des évaluateurs, implicite, non totalement conscient même si agi et agissant, pour ensuite le montrer, le modifier éventuellement, lutter contre de possibles dérives (critères oubliés ou surestimés...).

Cela peut aider les évaluateurs à prendre conscience de ce qu'ils font quand ils évaluent – une partie de l'activité restant tacite, « non consciente », comme c'est le cas dans toute activité de travail (Vermersch, 1994) – et de le discuter, de lui donner de la rigueur. Beaucoup d'études en sciences de l'éducation portent ainsi sur la question des biais de jugement et sur la manière de les prévenir (« biais de confirmation », « effet d'attente », poids des premiers jugements etc., voir notamment Merle, 2007).

Ceci permet aussi d'interroger l'activité des évaluateurs de programme, même si peu d'observations de situations de construction des jugements évaluatifs ont été réalisées. Les mêmes phénomènes de construction progressive du jugement sont-ils à l'œuvre ? Les évaluateurs de programme ont-ils toujours une pleine conscience des critères et référents effectivement pris en compte dans l'activité, ne souffrent-ils jamais de biais de jugement, comme le suggère par exemple Patton (1986) ? Hurteau, Houle et Mongiat (2009) ont par exemple montré, en analysant 40 rapports d'évaluation publiés dans des revues internationales, que peu d'évaluateurs explicitaient les « références » ou « standards » qui ont conduit la production de leurs jugements, et que même si les critères étaient la plupart du temps énoncés, ils restaient peu justifiés. Beaucoup reste à explorer sur la formation des jugements en évaluation de programmes, tout au long du processus, et sur tout ce qui les influence, dans la lignée des travaux plus récents de Hurteau, Houle et Guillemette (2012).

2.3. Étude de cas 2 : un évaluateur de politique publique au travail

À travers un deuxième court extrait de l'activité d'un évaluateur, nous allons illustrer les deux derniers phénomènes mis à jour, à savoir 3) qu'évaluer est une activité complexe faite d'adaptations et d'arbitrages et 4) qu'évaluer peut s'exercer de manière plus ou moins experte.

Cet extrait montre un responsable d'évaluation expérimenté en administration (travaillant dans une collectivité territoriale française) en train de concevoir un cahier des charges pour évaluer une politique publique locale d'aménagement urbain (pistes cyclables). Les données sont issues d'une thèse sur l'activité des évaluateurs de politiques publiques (Tourmen, 2007), et, plus particulièrement, de l'observation d'une activité provoquée – selon la méthode de « l'entretien clinique » inspirée des travaux de Piaget (Ghiglione et Matalon, 1992). Un cadre théorique de psychologie du travail et de didactique professionnel a été mobilisé (Pastré, Mayen et Vergnaud, 2006). Une étude de cas d'évaluation de programme – inspirée d'un cas réel - a été donnée à quatre professionnels, deux expérimentés et deux débutants, sur une durée deux heures, dans des conditions similaires. Sur la base d'une documentation écrite, il leur a été demandé de construire le cahier des charges de l'évaluation de la politique du vélo dans une agglomération. L'extrait ci-dessous est issu du discours du responsable d'évaluation qui commente son activité en même temps qu'il la réalise. Il a conçu des questions d'évaluation et tente à présent de les décliner en critères et indicateurs pour orienter la réponse à appel d'offres qui sera faite par des cabinets d'évaluation.

« Alors évidemment y'a dans le bilan des éléments intéressants qu'on peut exploiter qui concernent les difficultés, les nœuds, les endroits justement dangereux etc. Il est probable, mais c'est pas certain, que là on se trouve sur des points où justement il y a conflit entre plusieurs logiques, la logique routière, la logique transports en commun, et la logique de vélo. Bon, et c'est là où il faut se poser la question sur la façon dont les systèmes de décision,

en fait c'est comment se sont prises les décisions, et donc ça interroge sur le dispositif de pilotage. Je l'ai pas écrit parce que c'est volontaire, les questions évaluatives vont tomber chez les élus (rires), si je commence à poser des problèmes de pilotage, à dire que y'a des conflits entre les uns et les autres, euh... on risque de me rayer la question. Alors qu'elle est centrale. J'exagère un peu mais y'a une difficulté de cet ordre-là qui peut se poser. Donc je mets que c'est là (question) numéro 2, hein. Alors, exemple de critère. C'est conflit d'usage de la voirie. Vous voyez, là je suis sur des critères complexes... Conflits d'usage de la voirie dus à des décisions d'aménagement contradictoires. Euh je vais avoir... des indicateurs. (silence) Faut que je revienne au système de pilotage, en fait, où est-ce qu'il est? (Il le cherche dans les documents introductifs) Voilà. Comité technique... »

On constate tout d'abord que **l'activité évaluative est bien un « processus dynamique » et « contextuel »** (Hurteau, Houle et Guillemette, 2012, p. 87). L'activité de cet évaluateur est itérative, incertaine, faite d'allers et retours entre les informations disponibles et des hypothèses posées sur la situation (« *Il est probable, mais c'est pas certain, que...* »). L'évaluateur s'adapte : il pose un diagnostic sur la situation en prélevant plusieurs informations successives (« *y'a dans le bilan des éléments intéressants qu'on peut exploiter* », « *Faut que je revienne au système de pilotage* ») qui nourrissent une représentation de la situation évaluée. Cette représentation est provisoire, opératoire, on dira « fonctionnelle » en psychologie du travail (Leplat, 1985). En l'occurrence, il est en train de se construire une représentation de l'état de la politique, de son degré de réalisation et de ses « *nœuds* » ou « *problèmes* » ou « *manques* » pour pouvoir centrer le choix des questions, critères et indicateurs sur ces éléments – dans une finalité d'amélioration de ladite politique.

Cette adaptation est d'autant plus importante que les situations d'évaluation sont changeantes et « dynamiques »¹³, qui les rend non complètement prévisibles. Par exemple, certains élèves ont pu se mettre à pleurer lors d'une épreuve pratique d'évaluation (Tourmen *et al.*, 2013), ou un agenda politique être bousculé au plus grand désarroi des évaluateurs. Comme l'a dit un responsable d'évaluation en ministère dans la même recherche, « *Les évaluations sont des processus longs... donc il y a toujours le risque d'obtenir des résultats qui ne collent plus avec les préoccupations du moment. Parfois un sujet est mis à l'agenda alors que personne ne l'attend. Ce sont des choses qu'on ne peut pas contrôler* » (Tourmen, 2009b). L'adaptation est aussi anticipation, le diagnostic pronostic, comme le montre cet extrait : « *Il est probable, mais c'est pas certain, que (...) Bon, et c'est là où il faut (...) en fait c'est (...) et donc ça interroge sur (...) si je commence à (...) on risque de* ».

Un second phénomène apparaît clairement dans cet extrait : **l'évaluateur réalise de constants arbitrages entre plusieurs finalités**. Le professionnel réalise un compromis entre deux finalités, améliorer la politique (« *difficulté* », « *centrale* ») et faire accepter l'évaluation et ses résultats à venir par les élus (« *on risque de me rayer la question* »). Le professionnel est aussi guidé par des préoccupations de « *faisabilité* », peu visibles dans ce raisonnement-là mais davantage à d'autres moments de son activité (Tourmen, 2008, 2009b). Des compromis entre plusieurs préoccupations (utilité, acceptabilité, faisabilité) sont réalisés, ce qui rend les choix de l'évaluateur complexes. Ce ne sont pas uniquement des choix techniques mais aussi des choix politiques au sens d'une prise en compte du système d'acteurs qui aura à recevoir

¹³ Selon la définition donnée en psychologie du travail (Pastré, 1999) : une situation « dynamique » est une situation qui évolue en l'absence de l'intervention des opérateurs.

et utiliser l'évaluation. Plusieurs études ont ainsi montré que les évaluateurs étaient pris entre plusieurs finalités parfois contradictoires. Un jury d'épreuve pratique en lycée agricole est par exemple pris entre des exigences de rigueur de notation, de réalisation de l'activité et de maintien de la motivation des élèves (Tourmen *et al.*, 2013). Des évaluateurs de programme sont quant à eux tiraillés entre des objectifs de rapidité, de qualité du travail, de satisfaction du commanditaire, de rentabilité et un souci de l'intérêt général (Tourmen, 2008, 2009b). Les enseignants sont eux aussi pris entre des exigences de rigueur de la notation, d'apprentissage des élèves et de maintien de leur motivation qui peuvent donner lieu à des arbitrages différents selon les contextes. Comme le constate Merle, « *Le professeur n'évalue pas seulement des compétences, il doit aussi préserver des relations pédagogiques* » (2007, p. 8).

Cet extrait laisse finalement deviner **ce que l'expérience peut apporter aux évaluateurs** : on voit se dessiner, sous nos yeux, un raisonnement caractéristique des plus expérimentés : long, mettant en balance plusieurs finalités, se basant sur plusieurs indices de situation, posant des anticipations basées sur des savoirs d'expérience (« *si je commence à poser des problèmes de pilotage, à dire que y'a des conflits entre les uns et les autres, euh... on risque de me rayer la question. Alors qu'elle est centrale. J'exagère un peu mais y'a une difficulté de cet ordre-là qui peut se poser* »). Il a été montré que les évaluateurs s'adaptent différemment à cette variabilité des situations (Tourmen, 2008, 2009b), les débutants étant plus facilement désarçonnés que les plus expérimentés qui ont construit des règles d'action (règles de diagnostic et règles d'anticipation) pour s'y adapter (Chi, Glaser et Farr, 1988, Vergnaud, 1996).

Ces résultats interrogent la formation des évaluateurs : comment favoriser le développement des ressources issues de l'expérience chez les jeunes évaluateurs (facultés d'anticipation et d'adaptation aux variations de situation) ? Comment les préparer à la variabilité et l'imprévisibilité des contextes d'évaluation ? Comment favoriser la transmission des ressources issues de l'expérience sur les lieux de travail ?

Conclusion

Cet article n'a pas vocation à épuiser la richesse des études sur l'évaluation en éducation, mais à en présenter certains apports qui peuvent intéresser l'évaluation des programmes et des politiques publiques.

Nous avons vu que l'évaluation est un objet de préoccupation ancien en éducation, et que plusieurs auteurs ont produit des modélisations de l'activité évaluative, l'ambition de leurs modèles dépassant le contexte éducatif. Ces modèles ont été discutés et raffinés par des recherches portant sur l'observation d'évaluateurs au travail, qu'ils évaluent des apprentissages ou des programmes éducatifs. Nous avons tenté d'en synthétiser les apports en quatre points avant de les illustrer par des données issues de l'observation d'évaluateurs au travail, dans plusieurs contextes.

Ces modèles de l'activité évaluative et ces résultats de recherche sont-ils pour autant pertinents pour comprendre l'activité de ceux qui évaluent des programmes publics, en dehors du contexte éducatif ? Les premiers éléments de réponse semblent être positifs, même si peu d'études tentent à ce jour ce genre de comparaison. Des modèles plus fins et empiriquement testés permettant de décrire la logique de toute activité évaluative permettraient d'aller un cran plus loin dans l'ambition de Shadish, Cook et Leviton (1991), qui ont tenté de définir

à quoi ressemblerait une possible théorie de l'évaluation. Selon eux, « Une théorie idéale (jamais atteignable) de l'évaluation décrirait et expliquerait pourquoi certaines pratiques d'évaluation produisent certains types de résultats au regard des situations que les évaluateurs affrontent. Elle clarifierait (a) les activités, processus et buts des évaluations ; (b) expliquerait les relations entre les activités évaluatives et les processus et buts qu'elles favorisent ; et (c) ferait des tests empiriques pour identifier ceux qui sont en conflit avec la recherche ou certains savoirs en évaluation » (1991, p. 31). Et d'ajouter qu'une « théorie désigne un corps de savoirs qui organise, catégorise, décrit, prédit, explique et aide à comprendre et contrôler un sujet. Les théories le font de différentes façons, en cherchant des lois invariantes, utilisant des définitions et des axiomes pour déduire des propositions testables, et décrivant des processus causaux qui expliquent certaines relations » (1991, p. 30).

L'activité évaluative pourrait donc être non seulement mieux modélisée et mieux décrite, mais aussi théorisée à partir d'un va-et-vient entre modèles et tests empiriques. Pour avancer sur la connaissance de l'activité des évaluateurs en situation de travail, resterait à continuer à documenter une base de données empirique, dans différents contextes, ceci afin de continuer la comparaison et de mettre à jour des invariants et des spécificités dans l'activité évaluative. L'approche par l'analyse psychologique de l'activité, telle que réalisée en éducation, reste complémentaire d'autres approches sociologiques ou politistes qui mettent l'accent sur les contextes d'action, leur émergence et l'interaction entre les actes individuels et les logiques collectives. Une théorisation de l'activité évaluative, qui ne pourrait être que pluridisciplinaire, reste donc à construire.

Bibliographie

- ALLAL L., (2001). La métacognition en perspective In FIGARI G. et ACHOUCHE M. (Eds.), *L'activité évaluative réinterrogée*, De Boeck, Bruxelles.
- ALLAL L. et 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.
- ARDOINO J. et BERGER G., (1989). *D'une évaluation en miettes à une évaluation en actes* : le cas des universités. Matrice, Paris.
- ARDOINO J., (2001), De la méthodologie à l'épistémologie In FIGARI G. et ACHOUCHE M. (Eds.), *L'activité évaluative réinterrogée*, De Boeck, Bruxelles.
- BAMBERGER M.J., RUGH J. and MABRY L.S., (2006). *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Sage, Thousand Oaks.
- BARBIER J.M., (1985). *L'évaluation en formation*. Presses Universitaires de France, Paris.
- BASLE M.; (2008) ; *Suivi et évaluation des politiques publiques et des programmes. Economie, conseil et gestion publique*. Economica, Paris.
- BELAIR L., (1990). Évaluation : comparer en toute équité. Démarche de réflexion. *Mesure et Évaluation en Éducation* 13-2.
- BONBOIR A., (1972). *La docimologie*. Presses Universitaires de France, Paris.
- BONNIOL J.J. et VIAL M., (1997). *Les modèles de l'évaluation*, De Boeck, Bruxelles.
- BRESSOUX P. et PANSU P., (2003). *Quand les enseignants jugent leurs élèves*. Presses Universitaires de France, Paris.
- BONNIOL J.J., (1981). *Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires*. Thèse de Doctorat es Lettres et Sciences Humaines, Bordeaux.

- CARDINET J., (1986). *Évaluation scolaire et mesure*. De Boeck, Bruxelles.
- CHI T.H., GLASER R. et FARR M.J., (1988). *The Nature of Expertise*. Laurence Erlbaum associates, Hillsdale.
- CSE, (1996). *Petit guide de l'évaluation des politiques publiques*. La Documentation Française, Paris.
- DAUVISIS M.C., (2006). L'instrumentation en évaluation. *Mesure et Évaluation en Éducation* 29-1, 45-66.
- DE KETELE J.M., (1985). *Docimologie, introduction aux concepts et aux pratiques*. Cabay, Louvain La Neuve.
- DE KETELE J.M., (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et Évaluation en Éducation* 29-1, 99-118.
- DEMAILLY L., (2001). Évaluer les politiques éducatives. De Boeck, Bruxelles.
- DEMAILLY L., (2006). Évaluer les évaluations, In FIGARI G., MOTTIER LOPEZ L. (Eds.), *Recherche sur l'évaluation en éducation*, l'Harmattan, Paris.
- DEMEUSE M., (2006). Qu'indiquent les indicateurs en matière d'éducation ? In FIGARI G., MOTTIER LOPEZ L. (Eds.), *Recherche sur l'évaluation en éducation*, l'Harmattan, Paris.
- DONALDSON S. I. and LIPSEY M. W. (2008). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In SHAW I., GREENE J. and MARK M. (Eds.), *Handbook of evaluation*, Sage, London.
- FIGARI G., (1994). *Évaluer : Quel référentiel ?* De Boeck, Bruxelles.
- FIGARI G., (2006). Le référentiel entre théorie et méthodologie, In FIGARI G., MOTTIER LOPEZ L. (Eds.), *Recherche sur l'évaluation en éducation*, l'Harmattan, Paris.
- FIGARI G. et ACHOUCHE M., (2001). *L'activité évaluative réinterrogée*. De Boeck, Bruxelles.
- FIGARI G. et REMAUD D., (2014). *Méthodologie d'évaluation en éducation et en formation, ou l'enquête évaluative*. De Boeck, Bruxelles.
- FITZPATRICK J.L., (2004). Exemplars as Case Studies: Reflections on the Link between Theory, Practice and Context. *American Journal of Evaluation* 25, 541-559.
- FITZPATRICK J.L., CHRISTIE T.A. and MARK M.M., (2008). *Evaluation in Action: Interviews with Expert Evaluators*. Sage, Thousand Oaks.
- GHIGLIONE R. et MATALON B., (1992). *Les enquêtes sociologiques. Théorie et pratique*. Armand Colin, Paris.
- GUBA E.G. and LINCOLN Y.S., (1989). *Fourth Generation Evaluation*. Sage, Thousand Oaks.
- HURTEAU M., HOULE S. and MONGIAT S., (2009). *How Legitimate and Justified are Judgments in Program Evaluation?* *American Journal of Evaluation* 15, 307-318.
- HURTEAU M, HOULE S. et GUILLEMETTE F., (2012). *L'évaluation de programme axée sur le jugement crédible*. Presses de l'Université du Québec, Québec.
- JORRO A., (2006). Devenir ami critique. Avec quelles compétences et quels gestes professionnels ? *Mesure et Évaluation en Éducation* 29-1, 31-44.
- KUHN T., (1972). *La structure des révolutions scientifiques*. Flammarion, Paris.
- KUNDIN E.M., (2010). A Conceptual Framework for how Evaluators Make Everyday Practice Decisions. *American Journal of Evaluation* 31 (3), 347-362.
- LAVELLE J.M. and DONALDSON S.I., (2010). University Based Evaluation Training Programs in the United-States 1980-2008: an Empirical Examination. *American Journal of Evaluation* 31 (1), 9-23.
- LEPLAT J., (1985). Les représentations fonctionnelles dans le travail. *Psychologie française* 30-314, 269-275.
- MAYEN P., MAYEUX C. et SAVOYANT A. (2006). Construire la référence : une activité permanente des jurys de VAE In FIGARI G., RODRIGUES P., PALMIRA ALVES M. et VALOIS P. (Eds.) *Évaluation des compétences et apprentissages expérimentiels*, Admee Europe et Educa, Lisbonne.
- MERLE P., (2007). *Les notes. Secrets de fabrication*. Presses Universitaires de France, Paris.
- MOTTIER-LOPEZ L. et FIGARI G., (2012). *Modélisations de l'évaluation en éducation. Questionnements épistémologiques*. De Boeck, Bruxelles.
- MONNIER E., (1992). *Évaluations de l'action des pouvoirs publics*. Economica, Paris.
- PAQUAY L., DARRAS E. et SAUSSEZ F., (2001). Les représentations de l'autoévaluation, In FIGARI G. et

- ACHOUCHE M. (Eds.), *L'activité évaluative réinterrogée*, De Boeck, Bruxelles.
- PASTRÉ P., MAYEN P. et VERGNAUD G., (2006). Note de synthèse sur la didactique professionnelle. *Revue Française de Pédagogie* 154, 145-198.
- PATTON M.Q., (1986). *Utilization Focused Evaluation* 2nd ed. Sage, Thousand Oaks.
- PERRET B., (2001). *L'évaluation des politiques publiques*. La Découverte, Paris.
- PONS X., (2010). *Évaluer l'action éducative. Des professionnels en concurrence*. Presses Universitaires de France, Paris.
- RODRIGUES P., (2001). L'entretien de recherche comme pratique d'évaluation, *In* FIGARI G. et ACHOUCHE M. (Eds.), *L'activité évaluative réinterrogée*, De Boeck, Bruxelles.
- SCHWANDT T., (2005). The Centrality of Practice to Evaluation. *American Journal of Evaluation* 26, 95-105.
- SCHWANDT T., (2007). *The Role of Practical Knowledge in Learning*. Paper delivered at the 21st Annual Conference of the American Evaluation Association, "Learning to Evaluate, Evaluating to Learn", Baltimore.
- SCHWANDT T., (2008). The relevance of practical knowledge traditions to evaluation practice, *In* SMITH N.L. and BRANDON P.R. (Eds.), *Fundamental Issues in Evaluation*, The Guilford Press, New York, 29-40.
- SCRIVEN M.S., (1967). The Methodology of Evaluation. *AERA Monograph Series in Curriculum Evaluation* 1, 39-83, Rand McNally, Chicago.
- SCRIVEN. M.S., (1981). *The Logic of Evaluation*. Edgepress, Inverness.
- SHADISH W.R., COOK T.D. and LEVITON L.C. (1991). *Foundations of Program Evaluation Theories of Practice*. Sage publications, Newbury Park.
- STUFFLEBEAM D., (1980). *L'évaluation en éducation et la prise de décision*. NHP, Ottawa.
- TOURMEN C. et FIGARI G., (2006). La référentialisation : une façon de modéliser l'évaluation de programme, entre théorie et pratique. Vers une comparaison des approches en France et au Québec. *Mesure et Évaluation en Éducation* 29-3, 5-25.
- TOURMEN C., (2007). *Les compétences des évaluateurs : le cas des évaluateurs de politiques publiques*. Thèse de doctorat soutenue à l'Université de Grenoble II, 19 décembre 2007, disponible auprès de l'auteur sur demande.
- TOURMEN C. (2008). Les compétences des évaluateurs de politiques publiques. *Formation-Emploi* 194, 53-63.
- TOURMEN C., (2009a). L'activité évaluative et la construction progressive du jugement. *Les Dossiers des sciences de l'éducation* 22, 101-119.
- TOURMEN C., (2009b). Evaluators' Decision Making: Between Theory, Practice and Experience. *American Journal of Evaluation* 30 (1), 7-30.
- TOURMEN C. et DROYER N., (2011). *Une méthode pour analyser et outiller les jurys d'évaluation*. Actes du 23^e colloque de l'Admee-Europe, Paris, 12-14 janvier.
- TOURMEN C., DEBLAY S., LIPP A. et BOUDE A. (2013). *Évaluer en situation professionnelle*. Rapport de recherche pour la DGER, Ministère de l'Agriculture, de l'Alimentation et de la Pêche, Paris.
- TOURMEN C. et DEMAILLY L., (2012). Évaluations en tensions : pour une critique documentée et constructive de l'évaluation, *In* BARON G., MATIJAZIK N. (Eds.) *L'évaluation des politiques publiques. Défis d'une société en tension*. L'Harmattan, Société Française de l'Évaluation, Paris, 51-67.
- TYLER R.W., (1950). *Basic Principles of Curriculum and Instruction, Syllabus for Education* 360. University of Chicago Press, Chicago.
- VIAL M., (2001). Évaluation et régulation, *In* FIGARI G. et ACHOUCHE M., *L'activité évaluative réinterrogée*, De Boeck, Bruxelles.
- VERGNAUD G., (1996). Au fond de l'action, la conceptualisation, *In* BARBIER J.M. (Ed.), *Savoirs théoriques et savoirs d'action*. PUF, Paris.
- VERMERSCH P., (1994). *L'entretien d'explicitation en formation initiale et en formation continue*. ESF, Paris.