Presses de l'ADMEE

Christophe Gremion & Nathalie Younès
(sous la direction de)

Évaluation en éducation et formation
Ancrages et perspectives

Évaluation en éducation et formation
Ancrages et perspectives

Évaluation en éducation et formation
Ancrages et perspectives

Sous la direction de Christophe Gremion et Nathalie Younès

Portées par l'ADMEE-Europe et l'ADMEE-Canada, les Presses de l'ADMEE publient des ouvrages portant sur l'évaluation en éducation et formation. Leur vocation est de nourrir le débat et d'enrichir les pratiques dans la communauté internationale francophone des chercheurs et praticiens de ces domaines.

Les Presses de l'ADMEE publient des ouvrages scientifiques permettant de présenter les travaux de recherche issus des colloques internationaux, des activités des réseaux scientifiques thématiques, et plus largement de toute problématique portant sur l'évaluation en éducation et formation. Elles publient également des livres plus spécialement destinés aux praticiens, qu'il s'agisse d'analyse de pratiques, de conception d'outils et de dispositifs d'évaluation ou de formation.

www.presses-admee.com

© 2026 Les Presses de l'ADMEE
Éditeur : ADMEE Europe
c/o IRDP, Faubourg de l'Hôpital 43
Case postale 556, CH-2002 Neuchâtel, SUISSE
Impression : Libri Plureos GmbH, Friedensallee 273, 22763 Hambourg (Allemagne)

Conception couverture : Borsu Olivier - IFRES ULiège

ISBN : 978-2-9701-5366-5

Dépôt légal : Janvier 2026

A Gérard Figari

Sommaire

Introduction **13**

Chapitre 1. **40 ans, cela se fête !** **21**
Alain Bouvier 21

Chapitre 2. **MEE et e-JIREF : Parcours et enjeux de deux revues scientifiques francophones** **29**
Carla Barroso da Costa, Cathy Perret, Nathalie Younès 29

Chapitre 3. **L'éternel débat des notes à l'école** **45**
Charles Hadji 45

Chapitre 4. **Y a-t-il encore un avenir pour l'évaluation ?** **51**
Jean-Marie De Ketele 51

Chapitre 5. **L'évaluation formative et ses reconfigurations conceptuelles** **71**
Lucie Mottier Lopez 71

Chapitre 6. **Evolution, tensions et perspectives de l'évaluation pour toutes et tous** **89**
Françoise Pasche Gossin 89

Chapitre 7. **RCPE : des recherches collaboratives sur l'évaluation en éducation, d'une approche initialement marginale à une nouvelle alliance entre recherche et formation** **107**

Lucie Mottier Lopez, Céline Girardet 107

Chapitre 8. De l'évaluation comme objet d'étude à l'évaluation comme pratique de recherche 125
L'essor de la recherche-intervention évaluative, défis et perspectives 125

Lucie Aussel, Isabelle Bertolino, Dominique Broussal, Christelle Chauffriasse, Manon Delbreil, Pauline Hédacq 125

Chapitre 9. Biais du jugement évaluatif en contexte scolaire 153

Dylan Dachet, Ariane Baye, Dominique Lafontaine et Marc Demeuse 153

Chapitre 10. La mesure dans la langue de Molière 177
Etat des lieux et avenues à envisager 177

Dominique Casanova et Sébastien Béland 177

Chapitre 11. Compétences non académiques : malgré les difficultés de conceptualisation et de mesure, un intérêt majeur pour les chercheurs et les praticiens en éducation 201

Sophie Morlaix 201

Chapitre 12. La menace de l'évaluation : 40 ans de recherche 217

Fabrizio Butera 217

Chapitre 13. Comment « évaluation » et « didactique » s'articulent-elles dans deux champs de recherche anglophones majeurs : les Pedagogical Content Knowlegde et l'Assessment Literacy ? 231
Évaluation et didactique dans le champ des PCK et de l'AL 231

Céline Lepareur, Annick Fagnant et Raphaël Pasquini 231

Chapitre 14. Évaluer la professionnalisation **257**
Apport de 40 ans de publication de l'ADMEE 257
Christophe Gremion 257

Chapitre 15. Réseau thématique international Reconnaissance, valorisation et validation des acquis de l´expérience de l´ADMEE-Europe **277**
Dynamique d'un groupe de recherche et orientations épistémologiques, théoriques et empiriques des investigations 277
Carmen Cavaco et Pascal Lafont 277

Chapitre 16. Évaluer en situation de travail, un renouvellement des questions d'évaluation des compétences **293**
Patrick Mayen 293

Chapitre 17. Quand dire l'évaluation, c'est faire apprendre **305**
La performativité des discours sur l'évaluation 305
Marc Romainville 305

Chapitre 18. L'évolution du concept de feedback, au carrefour des dimensions cognitive, émotionnelle, comportementale et sociale de l'apprentissage **321**
Des dimensions inhérentes à l'engagement des apprenants dans le traitement du feedback 321
Matthieu Hausman, Pascal Detroz et Laurent Leduc 321

Chapitre 19. Culture, pratiques et évolution de l'évaluation en éducation et formation **339**
Quarante années de transformation et de recherche au Maroc 339
Ghizlane Chemsi, Naceur Achtaich, Mohamed Radid et Mohammed Talbi 339

Chapitre 20. Quarante ans d'évolution de l'évaluation au Liban : Enjeux, réformes et perspectives **357**
Wassim El-Khatib, Patricia Azoury 357

Chapitre 21. Du testing adaptatif à l'évaluation intelligente 375
40 ans de tissage entre le numérique et l'évaluation au sein de la Communauté ADMEE 375

Pierre-François Coen 375

Chapitre 22. Changer notre rapport aux preuves d'apprentissage : questions posées par les IA génératives en évaluation 393
Problématiser l'émergence de l'IA en évaluation des apprentissages en enseignement supérieur 393

Isabelle Nizet 393

Chapitre 23. 40 ans de crises de valeurs en évaluation 413
Crisologie de nos recherches 413

Alban Roblez 413

Liste des auteurs et autrices 439

Introduction

40 ans ! La durée d'une carrière scientifique. Une temporalité symbolique pour faire le point, se retourner sur la trajectoire de l'ADMÉE-Europe, pour revenir sur ses lignes de force et esquisser ses lignes d'horizon. Dans leur article paru à l'occasion des 30 ans de notre association, occasion d'un regard rétroprospectif sur ses origines, trois membres fondateurs, Linda Allal, Marie-Claire Dauvisis et Jean Marie De Ketele (2017) nous conviaient à poursuivre les réflexions en évoquant quelques pistes pour l'avenir. Ils pointaient notamment, l'internationalisation au niveau des membres et dans les publications, les synergies entre les différentes approches de recherche, une thématique-objet d'un ouvrage pour « nos » quarante ans.

Cet ouvrage collectif souhaite répondre à leur invitation et nous sommes particulièrement heureux que l'appel lancé par l'actuel président ait reçu un si large écho auprès des différentes générations, réseaux et tendances de recherches qui s'y expriment. Les contributions, rassemblées en deux parties, permettent de rendre visible comment les fils de la construction d'un socle de référence se nouent avec une dynamique d'élargissement. Chacune est constitutive de cette trame commune en devenir, une alliance entre retour rétrospectif, état des lieux et passage de témoin aux prochaines générations de chercheuses et de chercheurs.

L'ADMEE-Europe, pionnière dans la construction d'un socle de référence autour de l'évaluation

Le choix de l'évaluation comme cœur problématique des recherches et des pratiques, à un moment où les sciences de l'éducation prenaient place dans l'univers institutionnel, apparait aujourd'hui particulièrement pertinent alors que la demande dans ce domaine est croissante. Alain Bouvier nous rappelle certains éléments biographiques précieux pour en comprendre la genèse en suivant le trajet de liens entre des personnes partageant les mêmes objectifs de recherche sur les questions d'évaluation en éducation et qui, dès le départ, poursuivaient des collaborations internationales.

L'ADMEE-Europe, souvent considérée comme la petite sœur de l'ADMEE-Canada fondée neuf années auparavant, est née en 1986 pour

soutenir une dynamique de recherche émergente en Europe, et plus particulièrement en France, alors qu'elle était déjà institutionnalisée au Québec. Dans le contexte nord-américain, l'accent était mis sur la mesure, ce qui se manifestait tant dans la dénomination initiale de l'ADMEE-Canada (alors Association pour le Développement de la Mesure en éducation) que dans le titre de la revue de l'association (d'abord Mesure en Éducation, puis Mesure et Evaluation en Education). L'examen du processus de création de ces deux associations montre donc des trajectoires à la fois distinctes, entremêlées et s'influençant réciproquement, ainsi que le met en lumière le texte de Carla Barroso, Cathy Perret et Nathalie Younès dans leur analyse des parcours et des défis des deux revues scientifiques soutenues par l'ADMEE-Europe : MEE et E-JIREF.

Allal et al. (2017) retracent les spécificités contextuelles des trois pays fondateurs de l'ADMEE-Europe que sont la Belgique, la France et la Suisse. Alors qu'en Belgique, il s'agissait essentiellement d'universitaires, les participants français provenaient d'institutions universitaires et non universitaires alors qu'en Suisse, elles et ils étaient issus de centres de recherche hors université travaillant régulièrement avec une équipe de l'université de Genève.

Les différents chapitres de ce livre montrent à quel point les questions initiales s'avèrent transversales au cours du temps, et constitutives d'un socle de référence commun, de repères heuristiques interreliés, dialogiques, mobilisant quantitatif et qualitatif, sommatif et formatif, accompagnement et certification, normalisation et émancipation.

Ainsi, dans le champ de l'évaluation scolaire qui reste prédominant dans les publications de l'ADMEE-Europe, les travaux de Jean Cardinet demeurent largement cités, comme plusieurs auteurs et autrices le notent au fil de ces pages. Dans le numéro thématique de la revue e-JIREF (revue créée par l'ADMEE-Europe) qui lui a été dédié, Wirthner (2017) précise que ce spécialiste d'édumétrie mondialement reconnu dans la conception, la validation et l'utilisation des tests à l'aide des outils statistiques les plus avancés, notamment les modèles multiniveaux, a également prôné une approche méthodologique large alliant quantitatif et qualitatif, le tout, dans une orientation tant compréhensive que praxéologique. Également fondateur des travaux francophones sur l'évaluation formative (Allal et al., 1976), il a contribué à mettre en lumière le concept de régulation, s'avérant central dans la conception d'une évaluation source de changement et de progrès, tant individuels que collectifs (Cardinet et Laveault, 1996). Les défis de l'implantation réussie d'une évaluation soutien d'apprentissage (« assessment for learning » dans le monde

anglophone) dans les systèmes d'éducation, restent entiers (Laveault et Allal, 2016). En considérant les reconceptualisations de l'évaluation formative (élargie, soutien d'apprentissage, continue pour apprendre, inclusive), les textes de Lucie Mottier Lopez et Françoise Pasche Gossin, à rebours d'un discours dichotomique, amènent à prendre en compte de manière réfléchie les tensions fondamentales du champ de l'évaluation scolaire - entre uniformisation et différenciation, entre normalisation et émancipation, entre bienveillance et exigence, entre accompagnement et certification - pour fonder des pratiques conscientisées et améliorées dans une approche intégrative, éthique et humaniste.

Un autre fil conducteur porte sur la « référentialisation » ou méthodologie de l'évaluation partagée (Figari, 1994 ; Figari et Remaud, 2014) comme forme d'enquête évaluative. L'importance de trouver des terrains d'entente situés prend de plus en plus de place dans les recherches actuelles. Le propre de ce concept-méthode de référentialisation consiste à la fois à déterminer des référents dans le champ axiologique et théorique et, dans le contexte où l'évaluation se déroule, à impliquer les parties prenantes. Ce qui conduit aussi à construire des instruments d'analyse, dans lesquels les critères sont justifiés et les indicateurs élaborés collectivement pour en vérifier tant le degré de pertinence que celui de réalisation. Sur la base d'échanges préalables, de réflexions et de négociations, les partenaires peuvent ainsi se mettre d'accord sur le recueil et l'interprétation des informations ainsi que sur les actions consécutives à effectuer. Selon Jean-Marie De Ketele, cette orientation a ouvert le champ aux questions de valeurs, d'éthique et - en référence à Charles Hadji (2021) - à la dimension d'une évaluation à visage humain, à rebours d'une évaluation classante toujours dominante. Cependant, les synergies entre chercheurs et praticiens mais aussi entre théoriciennes et praticiennes, qui est un objectif majeur de l'association, restent un défi, bien que les développements des années 2000, et notamment la création du réseau « RCPE » (recherches collaboratives sur les pratiques évaluatives), marquent des progrès significatifs à ce propos, ainsi que Lucie Mottier Lopez et Céline Girardet le démontrent. Ce fil conducteur est également central dans la recherche-intervention évaluative, théorisée dans le chapitre rédigé par Lucie Aussel, Isabelle Bertolino, Dominique Broussal, Christelle Chauffriasse, Manon Delbreil et Pauline Hédacq qui mettent plus particulièrement l'accent sur la question du pouvoir et des processus démocratiques en jeu dans des démarches données à voir dans leur diversité et leurs exigences tant éthiques que méthodologiques.

Par ailleurs et depuis plus d'un siècle, la docimologie a accumulé une masse considérable de travaux montrant que le jugement évaluatif des

enseignants est affecté par des phénomènes inconscients (stéréotypes, attentes) qui désavantagent particulièrement les élèves discriminés (issus de milieux défavorisés, porteurs de handicaps…). Dylan Dachet, Ariane Baye, Dominique Lafontaine et Marc Demeuse offrent un retour bienvenu sur des études princeps et des méthodologies ayant étudié ces « biais » et des stratégies pour les atténuer - comme la correction anonyme et la focalisation sur une évaluation formative - afin de promouvoir un système éducatif plus équitable. Dominique Casanova et Sébastien Béland mettent en évidence les avancées récentes dans le domaine de la mesure, notamment l'utilisation des modèles de réseaux, des modèles cognitifs et de l'intelligence artificielle, pour améliorer l'évaluation et le diagnostic des apprentissages. Ils appellent à une réflexion sur les nouveaux standards de validité et à l'adoption de modélisations bayésiennes pour tenir compte des défis posés par les enquêtes à grande échelle. Enfin, Sophie Morlaix propose un état des lieux de la recherche sur l'évaluation des compétences non académiques, montrant d'une part leur importance croissante dans les parcours scolaires, universitaires et professionnels, et, d'autre part, les difficultés pour les conceptualiser et les mesurer.

Les élargissements de l'ADMEE-Europe

Il importe à l'occasion de cet anniversaire de relever à quel point la trajectoire de l'ADMEE-Europe a su s'élargir en s'ouvrant à d'autres champs d'intervention et d'enquête, d'autres cultures, d'autres pays, d'autres disciplines, d'autres approfondissements et renouvellements des problématisations.

Les textes qui composent l'ouvrage reflètent la pluralité de notre communauté. Des synergies interdisciplinaires se sont renforcées et enrichies en intégrant progressivement de nouveaux apports. Ainsi le chapitre de Fabrizio Butera présente des résultats issus de la psychologie sociale et qui irriguent désormais les travaux de l'ADMEE Europe. De leur côté, Céline Lepareur, Annick Fagnant et Raphaël Pasquini, en mobilisant deux champs de recherche anglophones majeurs (les « pedagogical content knowledge » et l'« assessment literacy »), mettent en exergue les articulations possibles entre la didactique et l'évaluation, s'inscrivant dans la lancée du réseau evadida, créé en 2013, en réponse au besoin de plus en plus vif d'établir des interactions.

A propos des domaines d'intervention et d'enquête, en plus de la centration initiale sur les élèves, on s'est intéressé aux professionnels de l'enseignement, de l'accompagnement et du pilotage du système scolaire. Par ailleurs, les travaux sur la formation professionnelle et les acquis de l'expérience ainsi que sur l'enseignement supérieur ont monté en puissance.

En examinant l'évolution du concept et des indicateurs de professionnalisation, Christophe Gremion met en relief les tensions entre le développement de l'autonomie individuelle et le contrôle institutionnel. De leur côté, en retraçant la dynamique du réseau international sur la reconnaissance, la valorisation et la validation des acquis de l'expérience (RVVAE) de l'ADMEE-Europe depuis sa création en 2003, Carmen Cavaco et Pascal Lafont soulignent la complexité de relier l'identification de savoirs expérientiels avec les enjeux éthiques de reconnaissance. Patrick Mayen explore pour sa part les enjeux de l'évaluation des compétences en situation de travail et plaide pour la mise en place de conditions institutionnelles sécurisées pour favoriser les apprentissages.

Deux chapitres, situés dans le champ de l'enseignement supérieur, portent également leur attention sur les effets des pratiques évaluatives sur les apprentissages des étudiants et étudiantes. Marc Romainville relève plus particulièrement l'impact considérable des discours enseignants et appelle à plus d'étude des pratiques ordinaires alors que Matthieu Hausman,

Pascal Detroz et Laurent Leduc s'intéressent aux conditions de l'efficacité des feedbacks, mettant l'accent sur la place de la personne étudiante et des médiations tant cognitives qu'affectives, sociales et comportementales, opérant dans le traitement du feedback.

Ces différents terrains sont autant de milieux spécifiques, montrant les apports et les limites des conceptualisations génériques ainsi que tous les chemins qui restent à parcourir ensemble.

Du point de vue des pays impliqués, le processus d'internationalisation originel belgo-franco-suisse s'est poursuivi avec l'ouverture des sections portugaise en 2002, luxembourgeoise en 2004, marocaine en 2009 puis libanaise en 2014. Les délégués fondateurs et actuels de la section marocaine, Ghizlane Chemsi, Naceur Achtaich, Mohamed Radid et Mohammed Talbi, ainsi que les délégués actuels de la section libanaise, Wassim El-Khatib et Patricia Azoury, offrent dans cet ouvrage les paysages de l'évolution de la recherche et des pratiques évaluatives dans ces deux contextes nationaux au regard des dimensions socioéconomiques et politiques qui les caractérisent.

Il semble que, depuis lors, ce processus d'internationalisation se soit stabilisé. Les contacts entrepris pour la création d'une section algérienne et d'une section tunisienne n'ont hélas pas encore abouti à ce jour, les crises politiques et économiques ayant certainement leur part dans cet état de fait. Faut-il y voir également un affaiblissement de la francophonie comme valeur porteuse ? Ou plus simplement, les limites des possibilités d'actions des enseignants-chercheurs et enseignantes-chercheuses assurant le portage de l'association, une mission qui s'ajoute, de manière bénévole, à toutes leurs autres missions chronophages d'enseignement, de recherche et d'administration ?

Enfin la technologie est désormais partie prenante de l'évaluation. Pierre-François Coen en retrace l'émergence, depuis le testing adaptatif ou test diagnostique informatisé des années 80-90 (relativement déceptif au regard de la complexité des phénomènes en jeu) qui trouve un regain d'intérêt grâce à la puissance des outils actuels jusqu'aux enjeux de l'IA qui semble aujourd'hui « écraser toutes les autres problématiques ». Cette technologie comporte des risques : risque de perte du sens de l'apprentissage, risque accru de standardisation des productions, de déshumanisation des interactions humaines ou encore de captation malveillante de données personnelles qui mettent le monde de l'évaluation au défi d'un usage éclairé, réfléchi et critique de l'IA pour une évaluation réellement intelligente ainsi que nous le propose Isabelle Nizet.

Si les recherches en évaluation sont aujourd'hui fortement ancrées dans les sciences de l'éducation, les origines disciplinaires extrêmement variées des chercheurs et chercheuses de ce domaine (mathématiques, didactiques, ingénierie, psychologie, lettres, économie, philosophie, sociologie, sciences de gestion, sciences cognitives…) et leurs différentes épistémologies, engendrent autant de spécificités dans l'étude scientifique des phénomènes évaluatifs. Ce pourrait être une explication aux foisonnements théoriques et méthodologiques caractérisant le champ ainsi qu'aux controverses le traversant. Il est à souligner que l'ADMEE-Europe réunit des communautés qui ne partagent pas nécessairement les mêmes paradigmes de l'évaluation, ce qui en montre la diversité, mais aussi l'importance de favoriser des débats, voire des controverses. La question des épistémologies est en effet en même temps transversale et ouverte, comme en témoignent les nombreuses publications établies à partir de la littérature internationale. Dès 1983, Chen et Rossi avaient insisté sur l'importance d'une « évaluation dirigée par la théorie », tant il est vrai que l'absence d'une telle conceptualisation renvoie les pratiques et recherches évaluatives à l'univers des croyances et des normes implicites. Les trois grandes strates paradigmatiques, quantitative, qualitative et participative qu'il est possible de déterminer à partir d'une étude cartographique des publications francophones et anglophones dédiées à l'épistémologie et aux épistémologies de l'évaluation (Younès, 2020), manifestent que ces différentes catégories de modèles coexistent mais aussi s'affrontent.

L'évaluation écologique, qui s'inscrit dans une perspective d'évaluation émancipatrice adaptative (Marcel et Gremion, 2025), invite à construire des méthodes et des dispositifs synergiques afin d'intégrer des complexités de différents types, en ménageant à la fois le niveau des situations locales, celui des dimensions transversales d'un commun validé, celui du sens de l'évaluation et de la manière de l'effectuer pour les personnes ou les collectifs évalués (Younès et al., 2020). Si les dimensions politiques et institutionnelles de l'évaluation en sont des composantes fondamentales, la part du vécu propre à chaque personne ou à chaque milieu évalué est également essentielle pour favoriser des dynamiques émancipatrices.

Mais le tout sans méconnaître que l'évaluation s'inscrit dans un monde en crise, marqué par l'incertitude et la désorientation. Le texte d'Alban Roblez pointe le caractère crucial du concept de crise dans l'évaluation : tension, point de bascule, effondrement du système apparaissent comme autant de situations critiques pour repenser des conditions de passage à un autre monde, un monde (plus) désirable.

I. Bibliographie

Allal, L., Cardinet, J. et Perrenoud, P. (1976). *L'évaluation formative dans un enseignement différencié*. Peter Lang.

Allal, L., Dauvisis, M-C. et De Ketele, J-M. (2017). L'ADMEE-Europe, née à Dijon en 1986 : développements et perspectives. *Evaluer. Journal international de Recherche en Education et Formation, 3*(3), 107-137).

Cardinet, J. et Laveault, D. (1996). Dix années de travaux européens et nord-américains sur l'évaluation : Quelles lignes de force ? *Mesure et évaluation en éducation, 18*(3), 1-25.

Chen, H.-T., et Rossi, P. H. (1983). Evaluating With Sense : The Theory-Driven Approach. *Evaluation Review, 7* (3), 283–302, https://doi.org/10.1177/0193841X8300700301

Figari, G. (1994). *Evaluer : quel référentiel ?* De Boeck.

Figari, G. et Remaud, D. (2014). *Méthodologie d'évaluation en éducation et formation, ou l'enquête évaluative*. De Boeck.

Hadji, C. (2021). *Le défi d'une évaluation à visage humain : dépasser les limites d'une société de la performance*. ESF Sciences humaines.

Laveault, D. et Allal, L. (dir.). (2016). *Assessment for learning : Meeting the challenge of implementation*. Springer.

Marcel, J.-F. et Gremion, C. (2025). *Évancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation.* Cépaduès Éditions.

Wirthner, M. (2017). Jean Cardinet : un précurseur. *Evaluer. Journal international de Recherche en Education et Formation, 3(*3), 13-21.

Younès, N., Gremion, C. et Sylvestre, E. (dir.). (2020). *Evaluations, sources de synergies*. Les Presses de l'ADMEE.

Younès, N. (2020). *Vers une évaluation écologique dans l'enseignement supérieur : dispositifs et transformations en jeu*. [Note de synthèse pour l'habilitation à diriger des recherches, Université de Lorraine].

Chapitre 1. 40 ans[1], cela se fête !

Alain Bouvier

En 1979, alors que je préparais ma thèse d'État[2] en mathématiques pures, je fus invité à faire un premier séjour de six mois à Queen's university, à Kingston dans l'Ontario ; d'autres allaient suivre. Outre ma spécialité de recherche, je m'intéressais à la didactique des mathématiques développée en France depuis la création des IREM en 1968, et à la formation des adultes peu qualifiés. En effet, j'avais eu la chance de travailler avec Bertrand Schwartz dans une longue recherche-action menée conjointement par les universités Lille 1 et Lyon 1 (où j'étais maître-assistant). Elle visait à créer et à assurer une formation diplômante (DEUG) ouverte à des adultes à Bac-3, pour les conduire, par unités capitalisables, en plusieurs années, à Bac+2, sans avoir à passer le Bac. Tout était individualisé et s'inspirait de la taxonomie d'objectifs de Bloom que Bertrand Schwartz faisait alors connaître en France[3]. Dans notre dispositif de formation d'adultes, l'évaluation jouait le rôle essentiel.

---

[1] Les 30 ans de l'ADMEE-Europe ont été présentés par les initiateurs du projet : Linda Allal, Marie-Claire Dauvisis et Jean-Marie De Ketele, dans un article sur lequel je m'appuie : L'ADMEE-Europe, née à Dijon en 1986 : développements et perspectives, disponible sur https://journal.admee.org/index.php/ejiref/issue/view/9

[2] Ce diplôme d'État, qui existait alors, est aujourd'hui remplacé par l'Habilitation à diriger des recherches (HDR).

[3] En 1971, membre du cabinet de Jacques Delors, il lui avait inspiré la loi sur la formation continue, connue sous le nom de « Loi Delors ».

Pendant mon premier séjour au Canada, outre les universités d'Ottawa et de Toronto, je suis fréquemment allé dans les universités québécoises pour participer à des colloques sur les sujets m'intéressant. C'est ainsi que j'ai entendu parler de l'ADMEE-Canada, très récente : créée en 1977, elle n'avait que deux ans d'existence. J'ai assisté à de riches rencontres et j'ai pu tisser des relations avec plusieurs participants. Certaines se poursuivent encore, je pense à Dany Lavaud de l'université d'Ottawa et Roland Louis de l'université de Sherbrooke, mais aussi Louise Bélair[4], présidente de l'ADMEE, qui assurait des cours en présentiel à Ottawa, suivis en direct sur des TV par des étudiants à Toronto. C'était ma première découverte de systèmes hybrides, bien avant Internet.

En France, à cette époque, et sauf quelques exceptions comme la docimologie et les travaux d'Henri Piéron, les questions d'évaluation étaient peu travaillées à l'université, c'est même un euphémisme ! Je fus surpris d'apprendre qu'au Québec, il en était tout autrement : des universités délivraient des diplômes professionnels en évaluation. Comme futur métier, l'évaluation attirait de nombreux étudiants, au point que cela finit par poser un problème social délicat : leur nombre dépassait de loin les possibilités d'emplois après l'université. Le système fut très vite saturé, si bien que la majorité de ces diplômes disparurent. Néanmoins, la culture de l'évaluation était installée dans le paysage universitaire et éducatif québécois.

Un peu plus tard, en 1984, de retour en France après plusieurs séjours en Amérique du Nord (Canada et USA), alors que je dirigeais l'IREM de Lyon, le chef de la MAFPEN[5] me proposa de participer à une formation-action nationale sur l'évaluation de politiques publiques, animée par François Pièrre, consultant et formateur d'adultes, lié au CNAM. À cette occasion je fis la connaissance d'un autre participant, Gérard Figari, aussi novice que moi[6], avec lequel j'avais sympathisé très vite. Le groupe en formation était chargé d'assurer, sous la direction méthodologique de François Pièrre, une commande officielle : fournir au ministère, avant une date précise, une

[4] Des années plus tard, devenu directeur de l'IUFM de Lyon que j'avais fondé, j'ai à deux reprises invité Louise Bélair à venir un trimestre pour faire des cours sur l'évaluation aux étudiants et aux formateurs de l'IUFM très peu formés à ce sujet ; c'était la situation en France. Je fis de même avec Roland Louis, autre membre de l'ADMEE-Canada.

[5] Structure alors responsable de la formation continue de tous les personnels de l'académie, soit plusieurs dizaines de milliers de personnes.

[6] Lui littéraire, moi mathématicien, nous étions faits pour nous entendre !

évaluation des 130 récentes universités d'été financées par l'État, afin qu'il puisse décider d'arrêter, de continuer à l'identique ou d'étendre ce dispositif original et très utile. En matière d'évaluation de politique publique, ce fut pour nous une grande leçon sur les rapports entre évaluateurs et commanditaires. Le travail de notre groupe se faisait le week-end et, à la fin, Gérard et moi fûmes chargés de la rédaction du document à remettre. Avec François Pièrre, à la date prévue, nous allâmes tous les trois au ministère présenter nos conclusions. Nous apprîmes alors que la décision était déjà prise depuis quelque temps ! Afin de récupérer d'importants moyens financiers, en deux ans, le ministère mit fin à ce dispositif malgré son intérêt notable. Souvent, pour les évaluations de politiques publiques, les commanditaires espèrent, au mieux, mais sans le dire, avoir une simple confirmation de leurs décisions empiriques, pour faire de la communication. Cette pratique est donc très éloignée des théories que l'on trouve dans les ouvrages spécialisés en évaluation ! Au moins, pour Gérard et moi, cela aura soudé notre amitié !

La France n'avait pas participé aux rencontres belgo-suisses sur l'évaluation qui se tenaient depuis 1977. Il faut dire que jusqu'au milieu des années 1980, le paysage français sur ce sujet était réduit, très épars et disparate. Les évaluations de politiques publiques commencèrent à se développer seulement au milieu des années 1980. Elles furent renforcées en 2000 par le vote de la Loi organique des lois de finances (LOLF)[7]. Les différents types d'évaluation, notamment pédagogiques, se multipliaient. S'engageaient désormais des travaux assurés tantôt par des chercheurs d'origine disciplinaire variée incluant les sciences de l'éducation[8], tantôt par des experts (français et étrangers), des associations et des consultants privés… payés très chers ! C'était l'époque, en France, de la pédagogie par objectifs et de l'évaluation formative.

Pour en revenir au paysage français en 1977, de mémoire, je cite les acteurs qui étaient les plus visibles : les psychologues Jean-Jacques Bonniol et René Amigues avec l'évaluation formatrice et leurs équipes à Aix-en-Provence dans des approches plutôt cognitives, le mouvement de la pédagogie institutionnelle à Paris VIII, avec Jacques Ardoino, Guy Berger, Georges Lapassade, Michel Lobrot, Jacques Guigou et la revue Pour, dans des approches psychosociologiques selon lesquelles tout est relation de pouvoir et de contre-pouvoir. Je note aussi la présence de Bertrand

[7] Ainsi fonctionne la France dans sa grande singularité.

[8] Elles avaient désormais leur place au Conseil national des universités (CNU) qui joue un rôle déterminant sur la qualification des enseignants-chercheurs français.

Schwartz et Guy Jobert, avec la revue Éducation permanente, qui firent connaître en France la taxonomie de Bloom ; à Lyon, le chimiste Maurice Chastrette et à Montpellier Françoise Cros, chimiste elle aussi, avec lesquels je collaborais aux côtés de Bertrand Schwartz. Il y avait encore, à Dijon, Marie-Claude Dauvisis à l'INRAP[9] et l'économiste Jean-Claude Escher qui dirigeait l'IREDU. À Lyon, on notait la présence du jeune Charles Hadji et du directeur du Cepec, Charles Delorme, (qui n'était pas universitaire), et encore, à Marseille, d'Yves Chevallard en didactique des mathématiques, de Jean-Marie Barbier au CNAM, à Paris, sans compter quelques individus isolés à Nantes, à Caen…, le plus souvent en sciences de l'éducation. Les séminaires des équipes faisaient la part belle à l'épistémologie, à Gaston Bachelard et à l'histoire des sciences.

Les pays francophones européens et le Québec avaient alors, en matière de recherche en évaluation, quelques longueurs d'avance sur la France, où les sciences de l'éducation tardivement nées en 1968 peinaient à se développer dans seulement une quinzaine des quatre-vingt-dix universités[10]. On peut noter que dans l'Hexagone, le chemin vers l'évaluation s'est fait en trois temps, sur trois décennies : dans les années 1970 par la formation continue des adultes, puis la formation continue des enseignants (IREM et MAFPEN) dans la décennie 1980 et enfin la formation initiale des enseignants (IUFM) dans les années 1990. Vint alors le temps des experts et spécialistes en évaluation, sollicités par les pays francophones d'Afrique, par le ministère de la coopération tant qu'il a existé, ensuite par le ministère des Affaires étrangères, par l'Unesco, par l'Europe pour des projets sur appel d'offres et par d'autres instances internationales. C'est ainsi, par exemple, que Jean-Marie De Ketele et moi fûmes sollicités par quelques-uns des différents organismes que je viens de citer et par des pays, notamment le Canada (Ottawa), pour évaluer le plus gros centre de recherche en éducation du Québec. De leur côté, les pays anglo-saxons pouvaient se passer de notre pays. Ils étaient riches de compétences et d'expertises, comme j'ai pu le constater lors d'un séminaire d'une semaine à Georgetown university (Washington DC) et un autre, aussi important,

[9] Dépendant du ministère de l'agriculture. L'enseignement agricole était tourné vers l'innovation et on peut noter que, de leur côté, les lycées professionnels étaient en pointe sur l'évaluation pédagogique, par exemple avec le Gérex soutien à la fin des années 1980.

[10] Elles sont encore absentes du CNRS (25 000 emplois, dont plus de 11 000 chercheurs) et du Collège de France.

regroupant des experts américains[11], mais en Colombie-Britannique : le travail se fait encore mieux dans un cadre magnifique et très agréable !

Je note aussi une particularité française de cette longue période fondatrice. Les origines disciplinaires des chercheurs en évaluation étaient variées : mathématiques, didactique des mathématiques ou de la physique, chimistes, École polytechnique et École des mines, CNAM, psychologie, lettres, psychosociologie, économie, philosophie, sociologie, sciences de gestion et, un peu, en sciences de l'éducation. Depuis, les temps ont changé.

C'est en 1986 que Linda Allal, Jean-Marie De Ketele et Yvan Tourneur[12], lors d'un colloque international à Dijon auquel je participais, soumirent aux débats des présents l'idée de créer en Europe une association internationale sur le modèle de l'ADMEE-Canada. Il fut proposé de l'appeler ADMEE-Europe. Mais dans l'acronyme canadien, le « M » était là pour « Mesure », idée inacceptable pour les institutionnalistes français et d'autres courants. Pour sortir du débat idéologique et faussement épistémologique, il fut alors décidé que le « M » serait conservé, mais au titre de la « Méthodologie » et Yvan Tourneur fut élu comme premier président. Plus tard, le Canada nous a rejoints sur cette appellation.

Dès la création de l'ADMEE-Europe, il fut décidé que la présidence de l'association tournerait tous les deux ans, en changeant de pays, ce qui se fait depuis. Et plutôt que de se doter d'une nouvelle revue en évaluation pour laquelle nous avions encore peu de forces vives en Europe francophone, si ce n'est autour de Jean Cardinet avec le renfort de Linda Allal et Yvan Tourneur, il fut décidé de proposer à l'ADMEE-Canada que leur revue devienne commune aux deux associations. Ce fut une très sage décision, du gagnant-gagnant !

Longtemps après arriva le moment où les Français devaient, à leur tour, proposer un candidat pour succéder à Jean-Marie De Ketele, président en exercice. Dans l'Hexagone, l'évaluation s'était développée, de petites équipes émergeaient ici et là, mais en se faisant concurrence, avec des désaccords surtout idéologiques, parfois violents, comme souvent en

[11] Outre un Allemand et un Français (ma personne), il regroupait des enseignants chercheurs et des experts américains plus politiques, dont deux membres du cabinet de deux différents Présidents américains, politiquement opposés, en présence de leur homologue en exercice auprès du Président du moment. Aux USA, à cette époque, c'était possible. En France, ce serait inimaginable !

[12] Cf. Note 1, op. cit.

France. Je crois que n'ayant pas d'enjeux personnels sur ce sujet et étant externe aux sciences de l'éducation, à mon grand étonnement, il me fut proposé d'assurer cette présidence. Je le fis avec plaisir. Cela me donna l'occasion de collaborer avec des gens que je trouvais fort sympathiques et compétents, en Suisse romande et en Wallonie, notamment dans l'équipe autour de Jean-Marie De Ketele avec le regretté Léopold Paquay qui devint mon vice-président. À Grenoble où était désormais Charles Hadji, Gérard Figari était déjà en pointe. Il sera plus tard le deuxième président français.

Au milieu des années 1980, il était curieux pour moi de constater que, hormis quelques pratiques de « pédagogie par objectifs » et de travaux sur les compétences de Guy Le Boterf, proche de Bertrand Schwartz, l'évaluation pédagogique était quasiment absente dans l'enseignement scolaire français, à l'exception de l'enseignement agricole toujours soucieux d'innovations et d'expérimentations, et de l'enseignement professionnel, en particulier avec le très intéressant « Gérex soutien[13] ». Impossible d'imaginer que l'enseignement général s'inspire de ces deux domaines qu'il méprisait par tradition, c'est la France !

Au début de l'ADMEE-Europe, j'avais noté que dans l'Hexagone, sur différents terrains autres que l'enseignement, comme le secteur médical par exemple, on voyait se pratiquer de l'expérimentation, du retour d'expériences et de l'évaluation. Il y avait des évaluateurs faisant réellement de l'évaluation. Dans notre secteur éducatif émergeaient surtout des auteurs de textes proposant, au niveau méta, des réflexions sur l'évaluation, souvent écrits par des chercheurs qui n'avaient pas, eux-mêmes, réalisé des évaluations. Comme s'il y avait d'un côté ceux qui faisaient empiriquement de l'évaluation, mais sans vues théoriques, et de l'autre, ceux qui théorisaient, mais sans expérience effective. J'imaginais alors l'ADMEE comme un lieu de riches rencontres entre ces deux mondes culturels qui s'ignoraient. Pendant les deux années de ma présidence, j'ai tenté, mais très naïvement et vainement, de faire évoluer cette situation, sans arriver à combler ce fossé. Heureusement, aujourd'hui, le temps a fait son œuvre. Les évolutions culturelles ne se décrètent pas, elles sont longues à aboutir.

[13] Dispositif pédagogique de soutien pour les élèves de lycées professionnels, fondé sur l'évaluation des compétences et conçu à la fin des années 1970 par de psychopédagogues des École normales nationales d'apprentissage, les ENNA chargées de la formation initiale et continue des professeurs de lycées professionnels jusqu'en 1990.

Avec son recul et sa grande sagesse, Charles Hadji nous dit, dans ce qui suit, où nous en sommes. Après l'absence d'évaluations, serait-ce le trop-plein ?

Chapitre 2. MEE et e-JIREF : Parcours et enjeux de deux revues scientifiques francophones

Carla Barroso da Costa, Cathy Perret, Nathalie Younès

I. Introduction

Les revues scientifiques jouent un rôle essentiel dans la diffusion des recherches francophones, la langue constituant à la fois un espace culturel de référence et une source de pertinence scientifique ainsi que d'intercompréhension. Comme le souligne Larivière (2014) dans une chronique publiée dans le Magazine de l'ACFAS, les revues jouent un rôle central dans la valorisation des travaux issus des équipes de recherche locales. Elles mettent en lumière la qualité des études menées et la contribution de la communauté scientifique à l'enrichissement des savoirs. En s'appuyant sur des processus éditoriaux rigoureux, elles alimentent les débats contemporains et soutiennent la prise de décisions éclairées (St-Onge et al., 2021).

Dans le champ de l'évaluation en éducation, l'ADMEE-Europe appuie deux revues scientifiques. Historiquement, Mesure et Évaluation en Éducation (MEE), fondée en 1978 par l'ADMEE-Canada, constitue un lien privilégié entre les deux associations. Plus récemment, Évaluer. Journal international de recherche en éducation et formation (e-JIREF), créé en 2015 par l'ADMEE-Europe, est venu élargir la politique de valorisation et de diffusion des savoirs scientifiques, tant pour les chercheuses et chercheurs que pour les praticiennes et praticiens engagés dans le domaine.

Un regard rétrospectif permet d'identifier les principaux domaines des articles publiés dans les deux revues : d'une part, la mesure, l'évaluation et la régulation des apprentissages en classe, dans l'enseignement supérieur et dans la formation professionnelle ; d'autre part, l'évaluation de programmes, de dispositifs et de systèmes éducatifs. Dans ces différents domaines, les questions relatives au numérique - et récemment à l'intelligence artificielle - aux dimensions tant affectives que sociales et cognitives, qui ont pris de l'importance ces dernières années, rendent compte des évolutions correspondantes du champ de la recherche et du champ des pratiques en évaluation.

Ce chapitre porte une attention particulière aux enjeux de la publication scientifique francophone, en mettant au centre de l'analyse l'expertise et les orientations de ces deux revues. Dans un contexte où l'anglicisation de la communication scientifique suit un processus difficilement réversible (Larivière et al., 2018) se pose la question de l'intérêt, pour la communauté francophone, de maintenir une politique éditoriale en français, malgré les efforts considérables qu'elle exige des équipes éditoriales comme des auteurs et autrices. L'expérience de MEE et d'e-JIREF, façonnée par cette tension constante, illustre à la fois des spécificités liées à leurs contextes tant culturels que nationaux et les stratégies mises en œuvre pour concilier rayonnement scientifique et vitalité de la publication francophone.

II. L'essor de MEE et d'e-JIREF dans la publication scientifique francophone en évaluation

II.a. Mesure et Évaluation en Éducation (1978) - « Le navire est lancé »[1]

C'est dans cette dynamique de valorisation du savoir scientifique francophone, au sein d'un champ alors en pleine émergence, que paraît, en mars 1978, le premier numéro du nommé bulletin Mesure en Éducation, un titre révélant le fort tropisme initial pour la mesure. Publiée par

[1] Extrait du numéro 1(2) de 1978 de la revue : introduction, écrit par Joseph Lemieux, premier directeur de la revue alors intitulée bulletin Mesure en Éducation.

l'Association professionnelle de mesure en éducation (APME), devenue, en 1979, l'Association des spécialistes de la mesure et de l'évaluation (ASPME), puis, en 1986, l'Association pour le développement de la mesure et de l'évaluation en éducation (ADMEE) et aujourd'hui, l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE), la revue, à l'origine, répondait à un besoin fondamental : maintenir et renforcer les liens de coopération entre les experts du domaine. En 1982, à l'occasion de la parution du numéro 5(1), la revue adopte son nom actuel : Mesure et Évaluation en Éducation (MEE).

À l'origine québécoise, la revue est née d'un double besoin : d'une part, promouvoir les échanges scientifiques et diffuser les résultats de recherche auprès des chercheuses, chercheurs et praticiennes, praticiens intéressés par un champ en développement ; d'autre part, offrir un espace de publication scientifique en langue française — une nécessité d'autant plus pressante dans le contexte de l'époque. En effet, comme le souligne le rapport de l'ACFAS (2021), les années 1970 et 1980 ont été marquées par la montée en puissance de l'anglais comme langue dominante de la production scientifique, souvent au détriment d'autres langues, dont le français. Dans ce contexte de marginalisation croissante des publications non anglophones, la revue MEE a su s'affirmer rapidement comme la principale revue francophone spécialisée en mesure et en évaluation, comblant ainsi un vide dans le paysage scientifique de l'époque.

En 1985, l'élargissement de la portée de la revue MEE s'est accéléré avec la création de l'association sœur, l'ADMEE-Europe, renforçant ainsi son caractère international. Le premier numéro marquant concrètement la composition d'une équipe de rédaction issue à la fois de l'ADMEE-Canada et de l'ADMEE-Europe est le volume 9(1), paru en 1986. Depuis lors, la collaboration entre les deux associations est une des forces de la revue, rassemblant une vaste communauté de recherche engagée dans la production d'un savoir scientifique rigoureux, ancré dans les contextes éducatifs francophones et enrichi par des contributions provenant de plusieurs continents. Cette assise s'est consolidée non seulement grâce à la rigueur de ses processus scientifiques et éditoriaux, mais aussi en raison de son approche résolument pluridisciplinaire. MEE se positionne ainsi comme un espace de publication ouvert à des domaines variés au-delà de l'éducation, tels que la santé et les sciences sociales.

II.b. e-JIREF (2015) – Naissance d'une revue aux multiples défis

En 2015, Évaluer. Journal international de recherche en éducation et formation (e-JIREF) publie son premier numéro et s'inscrit d'emblée dans le paysage de la recherche francophone en éducation et formation. Dès son éditorial fondateur, la revue affiche son ambition d'offrir « un regard pluriel sur l'évaluation, offert aux scientifiques francophones et construit par eux » (Demeuse et al., 2015, p. 1). Elle adopte une conception élargie de l'évaluation, dépassant les questions de mesure en éducation, pour intégrer les apports de disciplines variées et abordant des objets multiples : évaluation des apprentissages, des dispositifs, des politiques éducatives ou encore en formation professionnelle. Ce positionnement vise à croiser différentes perspectives méthodologiques et théoriques, afin d'enrichir la réflexion et d'inspirer les pratiques.

Complémentaire à MEE, e-JIREF s'inscrit dans l'écosystème de communication scientifique de l'ADMEE-Europe, aux côtés de ses colloques internationaux, de ses réseaux thématiques, de ses journées nationales et de sa newsletter. Le processus de publication se veut rapide, accessible et ouvert à toutes et tous, quelle que soit l'origine géographique des personnes autrices. Créée dans un contexte d'expansion des sections nationales de l'ADMEE-Europe — qui, après le trio initial Belgique-France-Suisse, a progressivement intégré le Luxembourg, le Portugal et le Maroc — la revue reflète la diversité et la richesse des travaux de recherche menés en français dans des environnements variés.

Respectueuse des standards scientifiques, e-JIREF revendique également une certaine liberté éditoriale, permettant d'exprimer les spécificités nationales et de traiter de problématiques parfois singulières. Fidèle à une vision démocratique de la diffusion des connaissances, elle rend les résultats de recherche accessibles non seulement à la communauté scientifique, mais aussi aux personnes engagées dans les pratiques éducatives et aux responsables impliqués dans le champ de l'éducation et de la formation.

III. L'accès ouvert pour la qualité et le partage démocratique des résultats de la recherche

Les trajectoires de MEE et d'e-JIREF les placent toutes deux parmi les acteurs engagés dans l'adaptation aux normes contemporaines de la publication scientifique. Elles défendent un modèle de diffusion en libre accès, refusant le système dans lequel les personnes chercheuses paient à la fois pour publier et pour lire. Elles contribuent ainsi activement à la démocratisation de l'accès au savoir en éducation et en formation.

Ce positionnement s'inscrit dans le contexte plus large de la mutation des environnements numériques et de la remise en question du poids croissant des grands éditeurs commerciaux. L'Initiative de Budapest pour l'accès ouvert, lancée en 2002, a constitué un jalon majeur en promouvant la mise en ligne gratuite, sans restriction, des articles évalués par les pairs, afin d'accélérer la recherche, enrichir l'enseignement et démocratiser la connaissance. Elle appelait aussi à repenser les mécanismes de financement, en mobilisant institutions, gouvernements et communautés savantes pour instaurer un modèle plus équitable et durable.

Au Canada et plus particulièrement au Québec, la transition vers le libre accès s'est accélérée à partir de 2015, avec l'adoption de la Politique des trois organismes[2] sur le libre accès aux publications, puis, en 2019, avec l'adhésion des Fonds de recherche du Québec (FRQ) à la cOAlition S et à son Plan S[3]. Depuis, le libre accès constitue un critère d'admissibilité incontournable aux programmes de soutien aux revues scientifiques, tant au niveau provincial (FRQSC) que fédéral (CRSH).

[2] Au Canada, les trois organismes subventionnaires fédéraux — le Conseil de recherches en sciences humaines (CRSH), le Conseil de recherches en sciences naturelles et en génie (CRSNG) et les Instituts de recherche en santé du Canada (IRSC) — souvent désignés sous le nom de Trois organismes, financent une grande partie de la recherche universitaire.

[3] cOAlition S, un consortium de grands organismes de financement européens, qui a lancé en 2018 le Plan S. Celui-ci stipule que toutes les publications scientifiques financées par des fonds publics doivent être accessibles immédiatement en libre accès, sans période d'embargo.

Entre 2021 et 2023, Mesure et Évaluation en Éducation (MEE) a consolidé son engagement envers le modèle « diamant »[4] (Coalition Publica, 2024), modèle également adopté par la revue e-JIREF. Les deux revues partagent ainsi une même vision : offrir un accès gratuit à leurs contenus, sans frais ni pour les personnes autrices ni pour les personnes lectrices, et promouvoir une diffusion équitable et ouverte des connaissances. Elles appliquent une licence libre (la licence Creative Commons CC BY 4.0) conforme aux standards du Plan S, et intègrent le code ORCID afin de garantir l'identification unique des auteur·e·s et de faciliter la reconnaissance de leurs contributions.

Ces choix illustrent la capacité de MEE et d'e-JIREF à s'adapter en continu aux standards contemporains de la publication scientifique, tout en partageant la volonté de rendre la recherche en éducation accessible, utile et librement réutilisable. Toutefois, comme le souligne le rapport de Beth et al. (2024), certains enjeux demeurent et trouvent un écho direct dans l'expérience concrète des deux revues.

IV. Mutations et adaptations au fil du temps : des enjeux en commun

Depuis le début des années 2000, le virage numérique des revues scientifiques a profondément transformé le paysage de la publication. La diffusion en ligne a modifié les modes d'accès au savoir, accéléré les processus éditoriaux, favorisé l'essor du libre accès et redéfini les critères de visibilité et d'évaluation des revues. Mais cette transition s'accompagne aussi d'effets plus critiques : pression accrue sur les équipes de rédaction pour maintenir des standards élevés dans des délais resserrés, risque de marginalisation des revues non anglophones, généralement moins bien indexées dans les grandes bases de données internationales, et besoin croissant de créer des espaces de partage, de discussion et de formation pour aider les équipes à s'adapter à ces mutations. Les enjeux liés à ces transformations seront présentés dans les sections qui suivent.

[4] Le libre accès diamant est un modèle de publication scientifique dans lequel ni les lecteurs ni les auteur·trices ne paient de frais. Ce modèle, non commercial et ancré dans la communauté académique, vise à garantir un accès gratuit et équitable aux savoirs. https://www.coalition-publi.ca/news-nouvelles/enonce-libre-acces-diamant

IV.a. Les enjeux de visibilité à l'ère du numérique

En conservant leurs caractéristiques, les deux revues partagent des atouts majeurs qui renforcent à la fois leur visibilité et leur crédibilité scientifique. Toutes deux sont indexées dans l'European Reference Index for the Humanities and Social Sciences (ERIH+), qui recense les revues répondant à des critères reconnus de qualité. Chacune bénéficie également d'une reconnaissance spécifique : e-JIREF et MEE sont référencées par le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES) en France, tandis que MEE figure également dans le Directory of Open Access Journals (DOAJ), gage de qualité éditoriale et de conformité aux standards internationaux du libre accès, ainsi que dans Education Source (EBSCO).

Actuellement, toutes deux utilisent la plateforme Open Journal Systems (OJS) pour la gestion des soumissions, ce qui, combiné à leur engagement commun pour la diffusion ouverte des savoirs, contribue au rayonnement de la production scientifique francophone en éducation.

Pour MEE, les transformations majeures s'inscrivent dans un contexte d'évolution constante des exigences en matière de diffusion scientifique. De 1978 à 2000, la revue publiait ses articles exclusivement au format papier, par le biais d'abonnements annuels individuels ou institutionnels. Le passage au numérique s'est fait progressivement : à l'automne 2000, elle propose à ses abonné·es un CD-ROM regroupant l'ensemble des articles déjà publiés — une première incursion dans le numérique, mais encore hors ligne. En 2014, MEE rejoint la plateforme du Consortium Érudit[5], offrant désormais ses articles à la fois en format papier et électronique. Ce choix stratégique joue un rôle déterminant dans l'accroissement de sa visibilité, Érudit occupant une place centrale dans la diffusion des connaissances scientifiques, particulièrement au Québec (Cameron-Pesant, 2018). En 2018, la revue franchit un cap décisif : elle abolit toute forme d'abonnement, lève l'embargo sur ses publications et adopte le libre accès intégral. Les articles sont dès lors diffusés exclusivement en format numérique, marquant la fin des éditions imprimées.

[5] Organisme à but non lucratif fondé en 1998 par l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal (UQAM), il a pour mission de soutenir l'édition et la diffusion en ligne des travaux scientifiques, en particulier dans les domaines des sciences humaines et sociales.

Pour sa part, E-JIREF, disponible en ligne dès sa création en 2015, a d'emblée adopté un modèle entièrement numérique, lui permettant de maîtriser sa diffusion et d'assurer un accès direct, rapide et gratuit à ses contenus. Ce choix lui a permis de se conformer aux principes du libre accès sans passer par la transition depuis un modèle papier.

IV.b. Les enjeux de la publication scientifique en français à l'heure de l'anglicisation

Un enjeu non négligeable réside dans l'intérêt croissant du milieu de la recherche à publier dans des revues anglophones et à fort facteur d'impact, ce qui leur permet de diffuser les résultats de leurs travaux auprès d'un auditoire plus large (Durand, 2010 ; Gingras et Mosbah-Natanson, 2010). Cette stratégie est souvent perçue comme gagnante, renforçant la visibilité des personnes chercheuses sur la scène internationale ainsi que leur reconnaissance institutionnelle, notamment en matière de financement de la recherche, de promotion académique et de positionnement dans les classements universitaires (Warren et Larivière, 2018). Cependant, cette dynamique exerce une pression croissante sur les communautés de recherche francophones, qui se voient incitées à publier en anglais, parfois au détriment de la production et de la valorisation du savoir scientifique en français. À titre d'exemple, une analyse de 14 280 publications en sciences humaines et sociales de l'Université de Montréal, couvrant la période de 1980 à 2019, révèle une baisse marquée de l'usage du français dans les publications : alors qu'il représentait plus de 50% des articles dans les années 1980, cette proportion est tombée à moins de 20% dans les années 2010 (Larivière, 2018).

Ainsi, les revues francophones, bien qu'essentielles pour diffuser des résultats ancrés dans les contextes socioculturels locaux, peinent souvent à rivaliser avec les grandes plateformes anglophones en matière de visibilité, d'indexation et de financement. Pour faire face à cette réalité, plusieurs stratégies émergent : certaines revues optent pour le bilinguisme, en acceptant la soumission des textes en français et en anglais, d'autres, comme est le cas de e-JIREF, choisissent d'assumer pleinement leur identité francophone en renforçant la légitimité de la publication en français, et d'autres comme la revue MEE, proposent des services de traduction vers l'anglais afin d'élargir leur lectorat et d'accroître la portée de leurs publications. Ainsi, depuis le volume 42 (2019), un numéro spécial intitulé Numéro Traduction/Translation Issue est publié annuellement, regroupant un nombre variable de textes traduits en anglais. La traduction

est proposée aux autrices et auteurs dont les articles ont été acceptés et publiés en français initialement et qui souhaitent bénéficier de ce service. Ce numéro additionnel est soumis à des conditions de publication qui peuvent varier d'une année à l'autre, en fonction des ressources financières disponibles.

IV.c. Une activité éditoriale exigeante reposant sur l'engagement des membres de la communauté

Comme le rappellent Beth et al. (2024), toutes les revues scientifiques, quel que soit leur domaine, subissent une pression importante pour maintenir des standards élevés en matière d'évaluation, d'édition et de publication, tout en respectant des délais jugés raisonnables et sans compromettre ni la rigueur scientifique, ni la créativité, ni la vitalité éditoriale. Ce défi est toutefois accentué pour les revues francophones, en particulier MEE et e-JIREF, dont le bassin d'autrices et d'auteurs, ainsi que de lectrices et lecteurs, demeure plus circonscrit au champ des méthodologies d'évaluation en éducation.

Dans ces deux cas, comme dans la majorité des revues, le travail éditorial repose largement sur un engagement bénévole, souvent invisible et peu reconnu au sein de la communauté universitaire (Larivière et al., 2021). Il s'agit d'une activité continue, exigeante en temps et en énergie, nécessitant une forte coordination pour assurer la qualité de toutes les étapes : gestion financière et logistique, coordination de l'évaluation par les pairs, planification des numéros réguliers et thématiques, préparation des demandes de subvention, résolution de problèmes techniques, etc. Tout cela s'ajoute aux autres responsabilités professorales, souvent sans allègement de tâche ni reconnaissance institutionnelle formelle (ACFAS, 2021). Dans ce contexte, les équipes de rédaction[6] doivent non seulement assumer ces tâches, mais aussi rendre visible l'ampleur de leur

[6] Par équipe de rédaction, nous désignons l'ensemble des personnes chargées d'administrer une revue, tant sur le plan financier et logistique que sur le plan éditorial. Leurs responsabilités couvrent notamment la supervision du processus de publication, la sélection des articles, le maintien de la qualité scientifique des articles, la définition des orientations et politiques éditoriales, ainsi que les activités de représentation et de mise en réseau.

contribution afin d'obtenir des dégrèvements[7] ou des soutiens adaptés, en faisant valoir l'importance stratégique de ces activités pour l'avancement des connaissances et la structuration des communautés scientifiques.

L'essor massif du numérique a ouvert de nouvelles perspectives, permettant de diversifier les formats et de donner à chaque revue une identité éditoriale distincte. Mais il a aussi accru les attentes, tant du côté des lecteurs et lectrices que des bailleurs de fonds. Dans les faits, les équipes apprennent souvent « sur le tas », adaptant leurs pratiques au fil des évolutions technologiques, faute de formations systématiques, celles-ci étant rares et souvent fragmentées. Des réseaux émergent néanmoins pour pallier ce manque. Par exemple, le regroupement des revues scientifiques de l'UQAM, le réseau Circé pour le soutien aux publications francophones au Québec et au Canada, ou encore Coalition Publica, qui œuvre à la construction d'une infrastructure de publication numérique. MEE est membre actif de tous ces réseaux.

De son côté, e-JIREF fonctionne grâce à un modèle qui garantit son indépendance scientifique, financé par les cotisations des membres de l'ADMEE-Europe et par les colloques annuels de l'association. Ces ressources rendues suffisantes grâce à l'engagement bénévole des membres de l'ADMEE-Europe, permettent à la revue de fonctionner et de se développer. Ceux-ci assument une diversité de rôles : rédaction, expertise, suivi éditorial, gestion technique de la plateforme, traduction de résumés, participation au comité scientifique, coordination de numéros thématiques, ou encore pilotage de projets spéciaux. Le lancement de la revue a été rendu possible par le soutien de collègues reconnus, qui ont accepté de publier dans ses premiers numéros pour lui conférer une crédibilité immédiate. Toutefois, cet équilibre reste fragile : surcharge des bénévoles, sollicitation répétée des mêmes personnes expertes pour l'évaluation, incidents techniques ou encore détection de plagiat nécessitant des réactions fermes et rapides de l'équipe.

V. Et quel est l'avenir des publications scientifiques ?

Les transformations rapides et constantes des environnements numériques, conjuguées à l'évolution des exigences en matière de

[7] Réduction de la charge d'enseignement.

publication, redéfinissent en profondeur le travail des revues scientifiques. Si l'avenir comporte encore de nombreuses incertitudes, certaines tendances se dessinent déjà clairement.

Parmi celles-ci, la recherche d'une transparence accrue dans le fonctionnement des revues, y compris sur le plan de la gestion financière, apparaît comme un enjeu central — particulièrement pour les revues francophones. À titre d'exemple, les Fonds de recherche du Québec (FRQ) attribuent désormais un identifiant numérique (Digital Object Identifier, DOI) à toutes les sources de financement, y compris aux revues scientifiques, afin d'assurer une meilleure traçabilité et visibilité des ressources octroyées et de leurs retombées concrètes.

Une autre tendance forte est la consolidation de réseaux d'appui aux équipes éditoriales. Ces réseaux, évoqués dans la section précédente, constituent des espaces de partage, de discussion et d'apprentissage qui prennent une place croissante dans les milieux universitaires francophones. Ils offrent des outils pour suivre l'évolution des normes de publication, mettent en lumière les conditions de travail des équipes éditoriales et incitent les institutions à mieux reconnaître leur contribution. Ils favorisent aussi la formation de la relève scientifique, en intégrant les personnes étudiantes universitaires aux activités des revues.

Enfin, la nécessité de dynamiser les pratiques éditoriales, sans compromettre la mission ni les standards de qualité, s'impose de plus en plus. Les revues explorent ainsi des projets novateurs, en phase avec les intérêts de la communauté scientifique et de leur lectorat, et adaptés à des contextes toujours plus inclusifs, numériques et internationaux. Dans ce mouvement de transformation, encourager et soutenir la participation des jeunes chercheuses et chercheurs constitue un levier stratégique essentiel pour assurer la vitalité et le renouvellement des revues francophones. Consciente de cet enjeu, l'ADMEE — au Canada comme en Europe — en a fait une priorité croissante, en développant des initiatives qui favorisent l'intégration de la relève scientifique, stimulent sa production et renforcent sa reconnaissance au sein de la communauté académique.

En Europe, plusieurs initiatives ont vu le jour pour faciliter l'entrée des jeunes dans la communauté scientifique. Lors de son congrès annuel, l'ADMEE-Europe décerne un prix à la meilleure communication présentée par une personne en début de carrière scientifique, accompagné de la possibilité de publier dans e-JIREF ou dans MEE, avec un soutien financier en cas d'acceptation, et toujours dans le respect des critères rigoureux d'évaluation. Cette action s'inscrit dans un dispositif plus large : ateliers sur la publication inaugurés au colloque de Casablanca en 2020, retraites

de rédaction organisées à Dijon en 2022 en partenariat avec l'IREDU, et ateliers « d'amis-critiques » lors du congrès à Mons en 2023 pour offrir une expertise bienveillante aux jeunes autrices et auteurs. En 2021, la création du concours à publication e-JIREF, destiné aux personnes étudiantes en master et en doctorat, est venue renforcer cette dynamique. En quatre ans, neuf lauréats — issus de France, de Belgique, de Suisse ou à double rattachement — ont vu leurs travaux publiés, reflétant la diversité des problématiques, des méthodes et des contextes explorés.

Au Canada, les sessions d'études annuelles de l'ADMEE-Canada offrent également des espaces pour soutenir la publication scientifique. Elles intègrent des discussions sur la revue MEE et des ateliers sur l'écriture scientifique, alignés sur ses politiques éditoriales. En 2023, un prix a été créé pour récompenser la qualité et l'originalité d'articles publiés par de jeunes chercheuses et chercheurs dans le volume 45 (2022). Cette mesure a eu un impact mesurable : la proportion d'articles issus de cette population est passée de 30% historiquement à 43% en 2023 et 2024, hausse attribuable à la promotion active de l'initiative lors des événements annuels de l'ADMEE au Canada et en Europe.

Qu'il s'agisse de l'ADMEE-Europe ou de l'ADMEE-Canada, ces actions convergent vers un même but : créer des passerelles concrètes pour encourager, accompagner et valoriser les contributions de la relève scientifique, tout en consolidant le dynamisme de la communauté et la pérennité des revues francophones.

VI. En guise de conclusion

Dans un contexte où l'anglicisation de la communication scientifique suit un processus qui paraît irréversible (Larivière et al., 2018), la question du maintien d'une politique éditoriale en français demeure centrale pour la communauté scientifique francophone. L'expérience de MEE et d'e-JIREF montre que ce choix, loin d'être anodin, relève d'un véritable positionnement stratégique et politique. Il s'agit non seulement de défendre la diversité linguistique, mais aussi de préserver un espace de production scientifique ancré dans des contextes locaux, pertinent pour les praticiens et praticiennes, et porteur de perspectives complémentaires à celles dominantes dans l'espace anglophone. Ces deux revues rappellent qu'une langue de publication est aussi un espace culturel et scientifique de référence, garantissant à la fois l'intercompréhension et la valorisation de

recherches souvent invisibilisées dans les circuits internationaux dominés par l'anglais.

L'avenir de la publication scientifique francophone repose ainsi sur un équilibre délicat : conjuguer exigence et ouverture, pérennité et innovation, ancrage local et rayonnement international. Les transformations liées au numérique, au libre accès et à l'essor de l'intelligence artificielle accentuent d'autant plus la nécessité d'adapter les pratiques éditoriales, de soutenir les équipes bénévoles et de reconnaître leur rôle stratégique dans l'écosystème scientifique.

Dans cette perspective, les normes et processus éditoriaux, tout comme les choix de formats, constituent des leviers déterminants qui orientent la trajectoire des revues et façonnent leur place dans le paysage académique. L'engagement de MEE et d'e-JIREF envers des standards élevés – assuré notamment par l'évaluation en double aveugle et l'utilisation de grilles critériées – confère rigueur et crédibilité aux contributions, tout en exposant les limites et biais propres à tout processus d'évaluation (Messu, 2020). Leurs orientations éditoriales, qu'il s'agisse de la liberté de style défendue par e-JIREF ou du format plus structuré privilégié par MEE, traduisent des visions singulières mais convergentes : offrir un espace de diffusion de savoirs pertinent, exigeant et accessible.

Au-delà des contraintes techniques et organisationnelles, ces revues incarnent une vision : celle de lieux où se rencontrent diversité méthodologique, pluralité des perspectives et ancrage dans des contextes éducatifs variés. Cette vision s'appuie sur un engagement affirmé envers la relève scientifique, considérée non pas comme un simple public cible, mais comme un acteur à part entière de la production et de la diffusion des connaissances.

Reste toutefois la nécessité de questionner certains choix éditoriaux qui peuvent contribuer à marginaliser certains formats ou à privilégier des modèles traditionnels de publication scientifique. Dans un environnement où la controverse constructive tend parfois à disparaître, il devient essentiel de rétablir un équilibre entre rigueur scientifique et ouverture aux débats argumentés, aux prises de position et aux analyses critiques. C'est à cette condition que les revues francophones pourront préserver leur vitalité intellectuelle et consolider leur rôle comme espaces vivants de production, de circulation et de confrontation des idées.

Pour les rédactrices et rédacteurs en chef qui prendront demain la responsabilité de MEE et d'e-JIREF, plusieurs éléments de réflexion méritent d'être retenus à la lumière des expériences déjà vécues. La

défense du français comme langue de publication scientifique constitue un point central : il s'agit d'assurer un ancrage solide en francophonie, tout en favorisant des collaborations nationales et internationales qui garantissent la visibilité des travaux dans un espace largement dominé par l'anglais. Il sera également essentiel de cultiver un haut niveau de rigueur éditoriale, en veillant à la transparence des processus et à l'équité des évaluations, sans pour autant étouffer la créativité ni la diversité méthodologique. Dans cette perspective, les formats publiés gagneront à demeurer de plus en plus ouverts et variés, afin de donner une place à des approches théoriques, empiriques, critiques, ou encore issues de contextes éducatifs peu représentés. L'innovation constituera un autre chantier incontournable, notamment par l'usage accru des outils numériques. Cette évolution devra toutefois être menée avec vigilance, afin de préserver l'intégrité scientifique, en s'appuyant sur des moyens adaptés et pertinents.

Enfin, la pérennité des revues reposera sur la force du collectif : reconnaître le travail des équipes bénévoles, répartir les responsabilités de manière durable et consolider des partenariats institutionnels solides. Bien évidemment, s'interroger sur le recours à l'IA par les autrices et auteurs comme par les expertes et experts. Être rédacteur ou rédactrice en chef ne consiste donc pas seulement à gérer un flux de publications, mais bien à faire vivre un espace intellectuel et culturel qui nourrit la réflexion, stimule le débat et soutient la vitalité de la recherche francophone.

VII. Bibliographie

Acfas, Institut canadien de recherche sur les minorités linguistiques, et Chaire de recherche du Canada sur les transformations de communication savante de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal. (2021, juin). *Portrait et défis de la recherche en français en contexte minoritaire au Canada* (rapport). https://www.acfas.ca/sites/default/files/documents_utiles/rapport_francophonie_final.pdf

Acfas (2021, juin). *Recherche et enseignement supérieur : leaders du rapprochement des francophonies canadiennes*. Mémoire produit dans le cadre du Sommet sur le rapprochement des francophonies canadiennes et de la prochaine politique du Québec en matière de francophonie canadienne.

https://www.acfas.ca/sites/default/files/documents_utiles/acfas_memoire_recherche_francophonie_2021.pdf

Beth, S., Henry, G., Fortier, A.-M. et van Bellen, S. (2024). *Reconnaître, valoriser, renforcer : recommandations issues du Symposium québécois des revues savantes*. Rapport Érudit et Acfas. https://erudit.org/public/documents/recommandations-symposium-revues.pdf

Cameron-Pesant, S. (2018). Usage et diffusion des revues savantes québécoises en sciences sociales et humaines : analyse des téléchargements de la plateforme Érudit. *Recherches sociographiques, 59*(3), 365–384. https://doi.org/10.7202/1058719ar

Coalition Publica. (2024, octobre). Énoncé sur le libre accès diamant. https://www.coalition-publi.ca/news-nouvelles/enonce-libre-acces-diamant

Demeuse, M., Fagnant, A. et Dupriez, V. (2015). Editorial. Un regard pluriel offert sur l'évaluation offert aux scientifiques francophones et construit par eux. *Evaluer. Journal international de Recherche en Education et Formation, 1*(1), 5-10.

Durand, C. X. (2006). If It's Not in English, It's Not Worth Reading! *Current Issues in Language Planning, 7*(1), 44–60. https://doi-org.proxy.bibliotheques.uqam.ca/10.2167/cilp085.0

Gingras, Y. et Mosbah-Natanson, S. (2010) Where are Social Sciences produced? Dans *World Social Science Report. Knowledge Divides* (p. 149-153), UNESCO Publishing et International Social Science Council. https://www.ost.uqam.ca/publications/where-are-social-sciences-produced/

Larivière, V. (2014, septembre). De l'importance des revues de recherche nationales. Chronique Les mesures de la recherche. *Magazine Acfas*. https://www.acfas.ca/publications/magazine/2014/09/l-importance-revues-recherche-nationales

Larivière, V. (2018). Le français, langue seconde ? De l'évolution des lieux et langues de publication des chercheurs au Québec, en France et en Allemagne. *Recherches sociographiques, 59*(3), 339-363. Érudit. https://doi.org/10.7202/1058718ar

Larivière, V., Beth, S., van Bellen, S., Delmas, E. et Paquin, É. (2021). Les revues savantes canadiennes en sciences humaines et sociales. Rapport Érudit. https://www.erudit.org/public/documents/Revues_canadiennes_shs_2021.pdf

Messu, M. (2020). Éthique et anonymat : Le cas des revues de sciences sociales françaises. Éthique publique. *Revue internationale d'éthique sociétale et gouvernementale, 22*(1). https://doi.org/10.4000/ethiquepublique.5423

St-Onge, S., Forgues, É., Larivière, V., Riddles, A. et Volkanova, V. (2021). *Portrait et défis de la recherche en français en contexte minoritaire au Canada*. Acfas. https://www.ost.uqam.ca/en/publications/portrait-et-defis-de-la-recherche-en-francais-en-contexte-minoritaire-au-canada/

Warren, J.-P. et Larivière, V. (2018). La diffusion des connaissances en langue française en sciences humaines et sociales. Les défis du nouvel environnement international. *Les revues savantes en sciences sociales et humanités, 59*(3), 327-333. https://doi.org/10.7202/1058717a

Chapitre 3. L'éternel débat des notes à l'école[1]

Charles Hadji

I. Introduction

Après des années de travaux sur l'évaluation des élèves et de si nombreuses recherches en « docimologie », on aurait pu croire la cause entendue : les notes attribuées au sein de l'institution scolaire ne sont ni une mesure objective des performances des élèves ni le moyen incontournable d'exprimer un jugement sur les niveaux atteints. Elles n'ont rien de scientifique. Et l'on pourrait, voire on devrait, s'en passer.

Mais la publication des résultats 2022 de l'enquête internationale PISA et la décision prise, dans la foulée, par le ministre Gabriel Attal, de décréter la fin du collège « uniforme » pour rendre le système éducatif français « plus exigeant », sont venues remettre la question des notes au cœur du débat éducatif.

Alors même que de nombreuses études dénoncent la pression excessive qu'elles exercent sur les élèves, l'ambition assumée d'une plus grande exigence va-t-elle leur redonner une place centrale dans l'institution scolaire ?

[1] Texte d'origine publié sur The Conversation France https://theconversation.com/profiles/charles-hadji-196837

II. Culture de la performance scolaire et bienveillance éducative

Depuis plus d'une décennie, le système scolaire semble être agité par deux mouvements pendulaires concomitants. Le premier axe de balancement concerne ce que l'on pourrait appeler sa philosophie dominante, qui oscille entre la proclamation de l'exigence, pour une recherche de l'excellence, et la volonté d'un accueil bienveillant de tous, dans un souci de réelle démocratisation.

Le second axe, qui recoupe le premier, concerne spécifiquement l'évaluation. Celle-ci est touchée par des poussées de fièvre périodiques dans le sens d'une pratique expansive et exacerbée de la notation, que l'on pense être au service de l'excellence, auxquelles succèdent des périodes d'accalmie, dans le sens d'une pratique moins invasive et plus apaisée – que l'on pense plus apte à contribuer à la démocratisation du système.

Ainsi, après que des voix, dénonçant une culture de la performance, et une obsession du classement, se soient élevées pour mettre fin à « la tyrannie de la note », le ministre Luc Chatel décide en novembre 2010 de maintenir la notation dans le système scolaire français. Car, déclare-t-il, la note est utile pour « avoir des repères ».

Puis, bien qu'un sondage IFOP ait révélé en 2012 que 80% des Français étaient opposés à la suppression des notes à l'école, on est tenté, en 2014, sur proposition du Conseil Supérieur des Programmes, de supprimer carrément notes et moyennes. Toutefois, en 2015, Le Monde titrera « La fin des notes n'est pas vraiment pour demain ». Une étude du CNRS montrera pourtant en 2016 que la suppression des notes en classe permet de réduire les inégalités de réussite liées à l'origine sociale (Huguet, 2016).

La question du brevet des collèges est symptomatique des débats accompagnant ce double mouvement. Depuis 2006, le brevet, qui vérifie la maîtrise des programmes, coexiste avec le Livret personnel de compétences, qui valide la maîtrise du socle commun de connaissances, de compétences et de culture. En 2014, le ministre Benoît Hamon envisage sa suppression, ou tout au moins, déjà, sa réforme. Doit-on conserver deux modalités d'examen ? Et quelle doit être la part d'une évaluation progressive, et celle d'une validation terminale ?

Le brevet a été maintenu jusqu'à aujourd'hui, avec une part de contrôle continu s'élevant à 50%. Mais Attal a fait part de son souhait de le réformer profondément, pour en faire un diplôme obligatoire pour entrer au lycée,

et qui ne prendra en compte que des notes données par des correcteurs, sans validation collective de compétences. Au risque de réduire les taux de réussite.

III. Evaluer pour accompagner ou pour classer ?

Le ministre de l'Éducation a donc décidé de rendre le système éducatif « plus exigeant ». Mais l'exigence (élitiste ?) exige-t-elle la notation, et la bienveillance (démocratisante ?) sa suppression ? Pour traiter sérieusement de la place et de l'avenir de la notation dans le système scolaire, il faut prendre acte du double fait que l'évaluation ne se réduit pas à la notation, et que la présence (ou l'absence) de la notation n'est pas, en soi, un signe d'excellence, pas plus que de médiocrité.

La note n'est qu'un système commode pour exprimer le jugement que l'on a pu porter sur la façon dont celui qui l'obtient satisfait à des attentes le concernant (en termes, par exemple, de construction de connaissances). Il existe d'autres façons, sans doute beaucoup plus informatrices, pour exprimer, et affiner, ce jugement – dont des outils de diagnostic personnalisé, telles les échelles descriptives. La présence de la notation n'a pas de signification univoque. Une note ne signifie pas rigueur, pas plus que son absence ne signifie laxisme ou faiblesse méthodologique. La rigueur tient dans le respect d'une démarche aujourd'hui clairement identifiée. Et la note n'est pas responsable de l'usage que l'on en fait !

Évaluer, c'est dire la valeur, en jugeant de l'acceptabilité d'une réalité par rapport à des attentes qui la concernent. Mais cela peut se faire dans deux grandes perspectives différentes : une perspective d'information et une perspective de filtrage. Dans le premier cas, l'objectif est de permettre à des individus de savoir, chacun, où ils en sont, dans le cadre d'un projet de progression individuelle. L'évaluation, alors informatrice, situe : elle répond à la question : où en suis-je ?

Dans le second cas, l'objectif est de dire la place d'un individu dans un ensemble où les places n'ont pas toutes la même valeur, dans le cadre d'un projet de progression sociale. Elle répond à la question : suis-je bien placé (voire, le mieux placé) pour monter dans l'« ascenseur social » ? L'évaluation, alors « classante », statue : elle permet de trier/filtrer ceux qui seront les mieux placés, parce que les mieux classés. Elle désigne les meilleurs, en leur attribuant, de fait, une plus grande valeur « académique » ; et, par là même, une plus grande valeur sociale.

IV. S'entendre sur les objectifs du système scolaire

Dans le second des cas évoqués, l'évaluation est conduite à privilégier le « couple infernal » classement/sélection, dans une logique, finalement, de concours. Dans le premier cas, elle tend à privilégier ce qui serait un « couple vertueux » : le diptyque diagnostic/accompagnement. On formule un diagnostic pour mieux accompagner chacun dans une progression individuelle. Il est clair que cette voie est préférable dans un contexte éducatif, quand il s'agit d'aider des élèves à apprendre. Et que la deuxième voie peut être, non préférable mais acceptable, si – et seulement si – le système scolaire reçoit pour mission légitime de réguler des flux en identifiant progressivement une élite.

Plutôt, donc, que de vouloir peser sur le mouvement de balancier pour donner une place plus ou moins importante à la notation, le mieux à faire serait de veiller à l'adéquation de la notation avec un objectif clairement défini. Mais après avoir pesé la légitimité de cet objectif, et en veillant à ne pas faire comme s'il n'y avait qu'un seul objectif, effaçant tous les autres.

Ainsi, il ne faut pas faire comme si l'évaluation ne pouvait être, par exemple, que « formative », en oubliant qu'elle doit avoir, aussi, pour fonction de certifier, en attribuant finalement des diplômes qui auront valeur de visa pour accéder à des positions socialement valorisées. Ou bien, comme si elle ne pouvait être que sélective, oubliant la nécessité, surtout à l'intérieur d'un système éducatif, de contribuer à la réussite de tous, en mettant en œuvre une évaluation formatrice.

Finalement, la place de l'évaluation (et pas simplement de la notation), dépend de l'importance et de la légitimité de l'objectif au service duquel elle est mise. On peut gager que, tant que le système scolaire aura pour fonction importante, voire, hélas, dominante, de filtrer une élite, l'évaluation par notation a un bel avenir devant elle. Car il faut bien nourrir l'ogre algorithmique (exemple : Parcoursup), qui a besoin d'informations classantes, aux points de passage clés du système scolaire, c'est-à-dire à chaque moment où une orientation décisive peut s'opérer. Mais il ne faudrait pas que l'évaluation ne soit réduite à ce seul service.

V. Bibliographie indicative

Collas, A., (2016, 10 mars). La suppression partielle des notes réduirait les inégalités entre élèves, *Le Monde*.

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

Hadji, C. (2015) (2018, nouvelle édition). *L'évaluation à l'école. Pour la réussite de tous les élèves.* Nathan.

Hadji, C. (2015, 3 octobre). Finissons-en avec la querelle de la notation. *Le Monde,* 15.

Hadji, C. (2016). Le système éducatif peut-il passer de l'évaluation normative à une évaluation constructive ? Dans B. Bocquet (dir.), *La fièvre de l'évaluation* (p. 113-135). Presses Universitaires du Septentrion.

Hadji, C. (2017). Faut-il avoir peur de l'évaluation en arts ? Dans D. Leduc et S. Béland (dir.), *Regards sur l'évaluation des apprentissages en arts à l'enseignement supérieur* (p. 147-165). Presses de l'Université du Québec.

Hadji, C. (2018). Ecole : l'évaluation, capable du meilleur, coupable du pire. *The Conversation, 17* octobre 2018. http://theconversation.com/ecole-levaluation-capable-du-meilleur-coupable-du-pire-104650

Hadji, C. (2019). Pour une évaluation humaniste, *Contextes et Didactiques* [En ligne], *13*. https://doi.org/10.4000/ced.811

Hadji, C. (2019). Débat : Comment faire bon usage des évaluations scolaires ? *The* Conversation, 6 mars 2019. http://theconversation.com/debat-comment-faire-bon-usage-des-evaluations-scolaires-112989

Hadji, C. (2020). École : évaluations nationales, est-ce le bon moment ? *The Conversation*, 13 septembre 2020. https://theconversation.com/ecole-evaluations-nationales-est-ce-le-bon-moment-145969

Hadji, C. (2021). *Le défi d'une évaluation à visage humain. Dépasser les limites de la société de performance*. ESF éditeur (Traduit en italien).

Hadji, C. (2022). Faut-il continuer à noter les élèves ? *The Conversation*, 19 septembre 2022. https://theconversation.com/faut-il-continuer-a-noter-les-eleves-184694

Hadji, C. (2023). Face à Parcoursup, le bac a-t-il encore une valeur ? The Conversation, 7 juin 2023. https://theconversation.com/face-a-parcoursup-le-bac-a-t-il-encore-une-valeur-207043

Huguet, P. (2016). Effets de la suppression partielle des notes sur les performances scolaires et les inégalités sociales. CNRS/Académie d'Orléans-Tours.

De Landsheere, G. (1974/1992). *Evaluation continue et examens. Précis de docimologie*. Labor/Nathan

Piéron, H. (1963*). Examens et docimologie*. PUF.

Chapitre 4. Y a-t-il encore un avenir pour l'évaluation ?

Jean-Marie De Ketele

I. Introduction

Un entrepreneur me faisait part récemment de son inquiétude au sujet de la multiplication des demandes d'évaluations. Et il ajoutait : « À quoi me servirait de savoir la quantité de produits sortis de mon atelier, si je n'ai guère d'idée de leur qualité pour les usages qui sont faits par leurs utilisateurs ? »

Cette réflexion a spontanément évoqué les nombreux écrits actuels qui ont dénoncé les effets pervers des pratiques évaluatives dans nos systèmes éducatifs. Et je transposais la question de cet ami entrepreneur comme suit : À quoi servirait-il de savoir que tel système éducatif ou de formation a sorti telle quantité d'élèves ou d'étudiants, si je n'ai guère d'idée de la qualité de leurs acquis pour leur devenir, par ailleurs marqué par l'incertitude ? Par facilité, les débats autour de l'évaluation se réduisent trop souvent à des positions idéologiques autour de l'opposition quantitatif-qualitatif et à des discussions techniques en termes de validité et de fiabilité. Ceci amène à négliger ce qui fait la qualité première de toute évaluation, la « pertinence ». Nous tenterons de montrer comment penser en ces termes permet d'éviter bien des démarches et des discussions inutiles, de se poser les questions essentielles, mais donc aussi de sortir des pratiques confortables, routinières et anonymisantes. Notre analyse montre enfin que l'avenir de l'évaluation passe par de nouveaux élargissements du champ couvert.

II. L'évaluation sous le feu des critiques

En 1959, Sorokin s'insurgeait contre la « quantophrénie », cette obsession de la mesure et de la quantification, c'est-à-dire cette tendance à traduire systématiquement les phénomènes sociaux et humains en langage mathématique, tendance qui engendre, selon lui, des effets de distorsion ou dérive : tendance à l'uniformisation, effets incitatifs inattendus et divers (Sorokin, 1959). Une telle critique a pris de l'ampleur ces derniers temps : Gori et al. (2011) dénoncent « La folie évaluation et les nouvelles fabriques de la servitude » ; Butera et al. (2011) considèrent « L'évaluation comme une menace » ; Hadji (2012) pose la question « Faut-il avoir peur de l'évaluation ? » ; Del Rey (2013), convaincue de la dangereuse inefficacité de « La tyrannie de l'évaluation », nous invite à résister en réinvestissant le paradigme de la complexité où l'imprévisible et la liberté ont toute leur place.

S'appuyant sur une recherche bibliographique menée avec les mots clés « évaluation » et « résistance », Lheureux et Jacob (2018) font l'inventaire des arguments de résistance à l'évaluation caractérisée par une perspective quantophrénique et une posture normative. Ces arguments sont nombreux. Plus fondamentalement, nous avons tenté de les catégoriser selon plusieurs dimensions :

- Au niveau macro, l'évaluation sert souvent une gouvernance soumise à une logique hiérarchique et légalo-comptable, dominée par un souci de contrôle de la performance et un besoin coercitif de surveiller et de parcelliser le travail (un néo-taylorisme).

- Au niveau de l'objet de l'évaluation, la quantification simplifie la réalité évaluée et ne tient pas compte de la singularité de la personne et de son contexte (son historicité et son environnement de vie). L'évaluation de ce qui est plus facilement mesurable pose la question de la pertinence (le rapport entre ce qui est évalué et ce qu'il faudrait évaluer) et la légitimité des critères et des indicateurs pris en considération.

- Au niveau des effets (souvent tacites) visés par l'évaluateur, l'évaluation est souvent une recherche et une manifestation de son pouvoir ; elle tend à justifier son besoin de classement et de hiérarchie des sujets ou des objets évalués ; et elle cherche à cacher la subjectivité du chercheur derrière des procédures mathématiques.

- Au niveau des effets observés sur les acteurs dont la mission est d'éduquer (enseignants, éducateurs, formateurs), l'évaluation provoque

un effet d'uniformisation du curriculum réellement mis en œuvre et évalué : cet effet a été particulièrement observé et mis en évidence en Angleterre, suite à la création de l'Ofsted[1] ; et on observe maintenant aussi les mêmes effets avec l'importance jouée par les évaluations nationales et internationales.

- Au niveau des effets perçus et observés auprès des personnes évaluées ou impliquées dans ce processus, l'évaluation provoque chez elles un besoin de se protéger contre les effets désagréables, comme le sentiment d'être jugé et contre l'anxiété liée aux conséquences éventuelles ; elle entraîne aussi des comportements de camouflage, de distorsion, voire des refus de fournir l'information. On touche ici le niveau éthique de l'évaluation.

- Plus fondamentalement, le recours intensif à la quantification dans l'évaluation reflète (ou masque) une épistémologie (néo-)positiviste : la réalité évaluée est réduite, morcelée, juxtaposée, sommée… et donne l'illusion de l'objectivité. Or la réalité est complexe. Son évaluation réclamerait donc une épistémologie de la complexité, comme le réclament les penseurs contemporains (Del Rey, 2013, mais aussi Morin avec ses cris d'alarme que sont *Réveillons-nous* de 2022 ou *Cheminer vers l'essentiel* de 2024).

Les arguments de résistance précités ont été le plus souvent énoncés par les auteurs de travaux sur l'évaluation aux niveaux micro (la classe) et méso (l'établissement). Il peut être aussi intéressant de consulter les experts-évaluateurs au niveau des politiques mondiales.

Spécialiste de l'évaluation au niveau macro, Adou (2017) met en évidence les limites des méthodes actuelles dans l'évaluation des ODD (Objectifs de Développement Durable, tels que définis et adoptés par l'Assemblée des Nations Unies en 2015). Il énonce quatre défis pour y répondre : (1) la coordination des acteurs impliqués ; (2) l'harmonisation des aspects méthodologiques, dont les critères communs ; (3) la qualité des données recueillies par les différents acteurs ; (4) le renforcement des capacités des acteurs impliqués par l'évaluation. Adou (2017, p.36) concluait « En effet,

[1] Voir l'article de Raveaud « L'inspection en Angleterre : les méthodes de l'Ofsted remises en cause », publié dans la Revue Internationale d'Éducation de Sèvres, 2024, n°97, 16-19. En plus des effets d'uniformisation, le suicide de la Proviseure de l'école Caversham, le 8 janvier 2023, suite à une inspection de l'Ofsted, a fait l'effet d'une bombe, non seulement en Angleterre mais bien au de-là, car elle a mis au jour l'insensibilité, la perversion et l'inhumanité potentielles de certaines pratiques d'évaluation.

toutes les approches utilisées actuellement pour les évaluations prises isolément ne peuvent répondre aux exigences de l'évaluation des ODD ». Ce constat peut être élargi à de nombreuses évaluations, non seulement macro, mais aussi meso (ex. : évaluation d'un établissement) et même micro (ex. : évaluation du suivi d'un élève en difficulté).

En septembre 2024, la Revue internationale d'éducation de Sèvres publiait un dossier sur « Les données en éducation », coordonné par Veran et Wagnon (2024a et b). L'entretien conduit par ceux-ci auprès du Secrétaire général de la CONFEMEN, Abdel Rahamane Baba-Moussa, ayant la responsabilité du PASEC (Programme d'analyse des systèmes éducatifs, l'équivalent de PISA pour l'Afrique francophone) confirmait les propos tenus par Adou (2017), en insistant sur la nécessité d'« Élaborer et diffuser des données fiables à tous les acteurs éducatifs et décisionnels en Afrique subsaharienne » (p.71). Mais quelles données produire ? Spécialiste international de « L'école de l'ombre » (le soutien scolaire privé payant), Bray (2024) s'étonne dans ce numéro du manque de données sur cet aspect dans les évaluations internationales et nationales alors que l'essor de ce mouvement va croissant et a des conséquences considérables : dépenses importantes des ménages ; exacerbation des inégalités sociales ; répercussion sur la scolarité régulière. Bray (2024, p.83) soulève donc le problème de la pertinence des données, c'est-à-dire le rapport « entre ce que l'on sait et ce qu'il faudrait savoir ». Dans leur étude de cas du Québec, Bachand et Demers (2024) insistent sur la nécessité d'une « culture critique des données » (p.99) : la pertinence des données non prises en compte dans l'évaluation et l'interprétation des données effectivement recueillies - notamment en regard de leurs finalités - soulève la question du rapport entre ce qui est réalisé et ce qui est souhaitable. À quoi servirait-il d'avoir des données parfaitement valides (le rapport entre ce qui est réellement mesuré et ce qui est déclaré) et parfaitement fiables (indépendance de la mesure par rapport aux évaluateurs et aux circonstances) si les données ne sont pas pertinentes. C'est pour répondre à de telles questions que Monteil (2024), dans son étude de cas sur la France, propose « une culture partagée de la production de données » (p.71) entre chercheurs et acteurs de terrain. C'est pour les mêmes raisons que Weinstein et Bravo (2024) rapportent un « nouveau système de collecte et d'usage des données pour l'amélioration scolaire durant la crise du Covid-19 au Chili » (p.113), palliant l'évaluation nationale rendue impossible pendant cette période, mais conservée par la suite, compte tenu des effets positifs générés. Les enseignants sont impliqués dans la production des données et les résultats sont immédiatement visibles. Enfin, Annan et Wootton (2024), dans l'expérience néo-zélandaise, vont

encore plus loin dans la volonté d'installer une culture partagée : les élèves sont mis au centre de la production et de l'analyse des données. Le scepticisme de départ de l'équipe éducative (chef d'établissement, enseignants, parents), qui pensait être la seule habilitée à fournir des commentaires aux élèves, « s'est mué en ravissement, à mesure qu'une dynamique naissait au sein de son équipe, qui s'enthousiasmait de la rapidité avec laquelle les élèves atteignaient leurs objectifs d'apprentissage » (p.135).

III. L'avenir est-il dans l'évaluation qualitative ?

Cela voudrait-il dire que l'avenir de l'évaluation serait entre les mains d'une évaluation plus qualitative ? Les expériences françaises, chiliennes et néo-zélandaises, énoncées ci-dessus, semblent plaider pour une évaluation davantage qualitative, sur la base d'un certain nombre d'arguments que nous pouvons catégoriser comme suit :

- Au niveau des acteurs : elle *permet* d'impliquer une pluralité d'acteurs, non simplement comme objets de l'évaluation, mais comme personnes qui ont un regard spécifique sur l'objet de l'évaluation ; la pluralité des regards permet de mieux cerner la complexité de l'objet et de la situation d'évaluation. *Encore faut-il* être en capacité de partager le pouvoir que donne l'évaluation !

- Au niveau de l'objet de l'évaluation : elle *permet* de prendre en considération de nombreux facteurs dont des aspects difficilement mesurables, de gagner en profondeur, de repérer des incidents critiques significatifs, de suivre des événements uniques ou inédits, de voir les évolutions des représentations et des comportements. *Encore faut-il* que l'évaluateur soit capable de (se) poser les bonnes questions (importance d'un bon cadre conceptuel et souplesse par rapport à celui-ci) !

- Au niveau des effets sur les acteurs : elle *permet* de prendre conscience que l'acte éducatif ou de formation, comme l'acte évaluatif, ne peut se réaliser pleinement que s'il y a engagement de tous les acteurs. Ceci induit de la modestie pour ceux qui sont en position de pouvoir et de la reconnaissance envers ceux qui ne le sont guère.

- Au niveau épistémologique : l'évaluation qualitative *permet* de cerner une réalité complexe : elle tente d'identifier les différentes dimensions de la réalité, de confronter les différents regards sur celle-ci, de dégager une décision partagée pour l'action. Ceci rejoint les fondements de

l'évaluation : celle-ci n'a pas pour fonction première le jugement, mais la prise de décision pour l'action ; le processus évaluatif implique des jugements divers, *mais* ceux-ci doivent servir une prise de décision pertinente pour l'action (De Ketele, 2013).

Ainsi donc, l'évaluation qualitative « *permet* » d'atteindre certains objectifs désirables, mais *encore faut-il* » que certaines conditions soient remplies. Et il est utile de confronter l'évaluation qualitative et l'évaluation quantitative en référence aux critères bien connus de pertinence, validité, fiabilité, discrimination et validité écologique (vs la généralisabilité tant désirée des chercheurs) :

- La **pertinence** exige que ce qui est évalué corresponde à ce qu'il faudrait évaluer. Dans l'évaluation quantitative de grande envergure, comme les évaluations internationales, le choix des objets repose en général sur un cadre conceptuel mûrement réfléchi et faisant l'objet d'un consensus assez large. La pertinence semble par là davantage garantie, sauf si le choix se restreint aux seuls objets facilement mesurables. Peut-on en dire autant dans les évaluations sommatives faites par les enseignants dans leur classe ? Certes, on peut dire que l'enseignant est censé être guidé par un référentiel de connaissances et de compétences qui a également fait l'objet d'un consensus assez large, mais son interprétation par l'enseignant peut être très subjective. Des auteurs comme Anadon et Guillemette (2007) tendent à considérer la subjectivité (et mieux, l'intersubjectivité) comme une valeur de l'évaluation qualitative, car dans des contextes précis, la pertinence de ce qui doit être évalué relève des acteurs présents, surtout dans les évaluations d'orientation et de régulation, mais même dans certaines évaluations de certification où les « capacités distinctives » (Dewey, 2011/1916) méritent d'être reconnues (Ricoeur, 2004).

- La **validité** exige que ce qui est évalué corresponde à ce qui est déclaré évalué. Une très grande attention est accordée à ce critère dans les évaluations quantitatives de grande envergure. Elles recourent à des opérations de pre-testing et à l'utilisation de modèles statistiques pour assurer leur légitimité. Il ne peut en être de même dans les évaluations scolaires, où il est souvent difficile de dire ce que signifie la somme des bonnes réponses à une série de questions : une performance certes, mais quelle performance ? La description qualitative des réponses données à ces questions dans tel contexte, historiquement situé, permet souvent de mieux nommer ce qui a été réellement évalué. Cet argument est à la base du développement des bulletins descriptifs, des portfolios ou des dossiers d'apprentissage (Scallon, 2000).

- La **fiabilité** exige que les résultats de l'évaluation soient indépendants de l'évaluateur et des circonstances de l'évaluation. Ce critère est l'objet d'une grande attention dans les évaluations quantitatives nationales ou internationales. En revanche, cette vérification est bien plus difficile à effectuer dans les évaluations scolaires sommatives, malgré les effets négatifs susceptibles d'être engendrés. Dans les évaluations qualitatives, où les différents acteurs sont activement impliqués, les désaccords exprimés constituent à la fois un objet d'évaluation et un matériau à partir duquel peut se construire une décision partagée, ouvrant la voie à une action éducative envisagée avec davantage d'assurance.

- Le **caractère discriminatif** est souvent revendiqué comme une qualité de l'évaluation quantitative : la mesure permettrait facilement de distinguer les acquis ou performances des différentes unités d'observation (les élèves, les enseignants, les établissements, les pays, les régions, …, selon le cas). Ici encore, des outils statistiques existent et sont utilisés dans les évaluations de grande envergure. Dans les évaluations sommatives au sein des établissements scolaires, une note de 10/20 obtenue grâce à la réussite de 10 items réussis ne signifie pas nécessairement que deux élèves ayant la même note aient les mêmes acquis. En revanche, deux notes différentes sont bien le fruit de deux performances différentes (mais lesquelles ? sur une même dimension ? sur des dimensions différentes, ramenées par un artefact mathématique à une dimension censée être commune ?). Ce faisant, la note porte un jugement sur la personne ou l'institution évaluée. Elle devient socialement discriminative et source d'injustice. L'évaluation qualitative descriptive, surtout lorsqu'elle est précise et qu'elle porte sur de nombreuses dimensions, discrimine par nature car il n'existe pas deux descriptions semblables. Il s'agit plutôt du caractère diversifié de l'évaluation : elle ne porte pas de jugement sur la personne ou l'institution si elle reste descriptive (des comportements, des faits, des éléments de contexte) et évite les appréciations (se méfier des adjectifs) qui ne sont pas le fruit d'une réflexion commune.

- **La validité écologique** exige que la procédure et les résultats de l'évaluation soient adéquats aux caractéristiques du contexte. C'est une qualité revendiquée par l'évaluation qualitative, à juste raison quand elle implique les différents acteurs dans l'évaluation. La validité écologique est nettement plus difficile à respecter dans les évaluations d'envergure qui reposent sur un même cadre conceptuel et une même procédure. Ainsi, les objets de l'évaluation n'ont pas la même pertinence dans tous les contextes ; un même indicateur ne valide pas nécessairement la même variable selon le contexte ; les évaluateurs de pays différents n'ont pas nécessairement les mêmes compétences pour évaluer, ce qui pose le

problème de la qualité des données[2]. La validité écologique serait-elle plus facile à respecter dans l'évaluation scolaire sommative au sein de la classe (même environnement spatial et de ressources) ? Les travaux sociologiques sur les inégalités scolaires[3] montrent que l'évaluation scolaire en est la principale cause, car l'environnement social et l'histoire de l'élève ne sont pas pris en compte *ou plutôt agissent tacitement* à travers l'évaluation.

IV. Un besoin d'élargir le champ de l'évaluation

Les expressions, telles « l'évaluation permet … encore faut-il », mettraient-elles en cause l'avenir de l'évaluation ? Cette question n'est pas neuve. En 1986, Cardinet nous la posait déjà dans deux ouvrages majeurs : « Évaluation scolaire et mesure » (1986a) et « Pour apprécier le travail des élèves » (1986b). Après une carrière de psychométricien aux États-Unis auprès de Thurston et Cronbach, devenu ensuite Directeur de l'Institut romand de recherche et de documentation pédagogique (IRDP) de Neuchâtel, il s'est attaqué « aux innombrables remarques, critiques et suggestions formulées à propos des systèmes d'évaluation » (Cardinet, 1986b, « Avant-propos ») qui lui venaient de toute part, non seulement dans le cadre de sa fonction mais aussi dans ses contacts avec les experts qui s'associaient dans la création et le développement de l'ADMEE[4]. Pour mettre de l'ordre dans le champ de l'évaluation, Cardinet (1986b) proposait d'élargir le champ (alors trop souvent réduit à la note scolaire et à des problèmes de docimologie) en se posant sept questions fondamentales interreliées : pour quoi évaluer ? pour qui évaluer ? sur

[2] L'entretien mené par Veran et Wagnon auprès du Secrétaire général de la CONFEMEN à propos du PASEC met bien en évidence cet aspect (Veran et Wagnon, 2024b). Ceci est souligné aussi par Adou (2017) à propos de l'évaluation des ODD.

[3] Voir par exemple la synthèse de Felouzis et al., (2013). Les marchés scolaires. Sociologie d'une politique publique d'éducation. PUF

[4] On ne rappellera jamais assez le rôle que Jean Cardinet a joué dans le champ de l'évaluation et dans la création de l'Association pour le Développement des Méthodologies de l'Évaluation en Éducation (ADMEE). Ce fut pour beaucoup d'entre nous un « maître », non seulement pour ses compétences dans le domaine, mais aussi par sa personnalité rendant l'Autre intelligent.

quoi évaluer ? quand évaluer ? comment recueillir l'information ? comment interpréter l'information ? comment utiliser l'information ?

Ces questions ont été à la base de travaux ultérieurs importants. Ainsi, le concept d'évaluation formative (initié par Allal, Cardinet et Perrenoud, 1976, puis développé et approfondi par Allal et Mottier Lopez, 2005) et celui d'évaluation formatrice (Nunziati, 1988) ont inspiré non seulement de nombreuses recherches, mais aussi les pratiques de terrain. De même, le concept de référentialisation de Figari (1994) a mis en évidence l'importance du pour quoi et pour qui l'évaluation. Ceci a soulevé la question de l'éthique (Gillet, 1991) et des valeurs dans l'éducation et l'évaluation (Reboul, 1992), et, plus largement, le besoin d'une évaluation plus humaine (Hadji, 1997). En passant en revue les différents paradigmes de l'évaluation, De Ketele (1993) montrait que chacun d'entre eux mettait l'accent sur des fonctions différentes (les pour quoi ? mais aussi le quand ?) et des procédures différentes (le comment ?), ce qui l'a poussé en 2013 à clarifier ces aspects en distinguant et en croisant les trois fonctions principales (orientation, régulation et certification) et les trois procédures (sommative, descriptive et interprétative) de l'évaluation. Les travaux de Vial ont surtout montré le besoin de se former à l'évaluation, ce qui implique de se donner une problématique et d'élaborer les concepts pertinents (Vial, 2009).

Correspondant à la naissance et au développement de l'ADMEE, ces travaux fondateurs se sont essentiellement centrés sur l'évaluation des apprentissages scolaires (« l'appréciation du travail des élèves », dit Cardinet), mais ont progressivement pénétré d'autres secteurs. En effet, si l'évaluation est essentielle dans l'apprentissage conçu comme transformation des acquis des élèves dans un environnement donné (*premier mouvement d'élargissement*), elle devrait l'être aussi dans la transformation des acquis d'autres acteurs personnels et institutionnels. D'autres mouvements d'élargissement ont donc émergé.

Tout naturellement, le *second mouvement* s'est tourné vers les professionnels de l'enseignement. Les travaux de Jorro (« Évaluation et développement professionnel », 2007) ont montré que de tels acteurs devaient être considérés aussi comme des sujets apprenants et en transformation grâce à un travail réflexif reposant sur un recueil descriptif de données et une démarche d'analyse, autant que possible avec d'autres acteurs, pour leur donner du sens et prendre des décisions afin d'améliorer

l'action. Sur la base de ce travail, trois symposiums du réseau REF[5], réunissant chaque fois une dizaine de chercheurs de plusieurs pays, ont approfondi des composantes du développement professionnel : « L'engagement professionnel en éducation et formation » (Jorro et De Ketele, 2013) ; « Mutations éducatives et engagement professionnel » (Merhan et al., 2015) ; « Les apprentissages professionnels accompagnés » (Jorro et al., 2017). Il est important de souligner que ce deuxième mouvement ne doit pas être pensé indépendamment du premier, car le développement professionnel de ces acteurs sert le travail de transformation des acquis des élèves. Les travaux ultérieurs pourraient évaluer davantage de tels impacts potentiels.

Logiquement, un *troisième mouvement* d'élargissement de l'évaluation semble naître. L'OCDE (2012) l'appelle de ses vœux quand elle parle de « Transformer les écoles en *organisations apprenantes* » ; Fernagu Oudet (2012) souhaite que celles-ci mettent en place en leur sein des « *environnements capacitants* ». Il ne s'agit pas de céder à la tentation quantophrénique des palmarès des écoles, mais de mener des travaux d'évaluation portant sur les caractéristiques de ces organisations quand elles deviennent apprenantes grâce à des environnements capacitants ; il s'agit aussi de montrer leur impact sur le développement professionnel de ses acteurs et indirectement sur la transformation des acquis des élèves. Ici encore, la démarche évaluative consiste à associer recueil descriptif des données pertinentes et analyse réflexive impliquant les différents acteurs de l'institution. Deux ouvrages traduisent bien l'esprit de ce mouvement : « L'évaluation, source de synergies » (Younès et al. (2020) et « Évancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation » (Marcel et Gremion, 2025)[6].

En faisant encore un pas vers le niveau macro, le besoin d'élargissement du champ de l'évaluation (le *quatrième mouvement*) porte sur la gouvernance au niveau d'une région ou d'un pays. Il ne s'agit pas simplement de la construction d'instruments d'évaluation des performances à ce niveau (ceci a déjà été fait ; et on en connaît les effets pervers), mais d'une véritable gouvernance où « les questions éducatives

[5] Le Réseau Éducation et Formation organise tous les deux ans des symposiums sur des problématiques de recherche contemporaines.

[6] Cet ouvrage implique dix chercheurs dans le champ de l'évaluation : outre les deux coordonnateurs de l'ouvrage, Charles Hadji, François Galichet, Lucie Mottier Lopez, Martine Janner Raimondi, Dominique Broussal, Alain Roblez, Nathalie Younès et Séphora Boucenna.

nécessitent de la part des cadres scolaires une maîtrise de plus en plus fine des *capacités analytiques, réflexives et argumentatives* » (Pelletier, 2009). Ici encore, on a besoin de travaux évaluant les effets de cette gouvernance sur les institutions scolaires (deviennent-elles des organisations apprenantes ?), sur les professionnels de l'enseignement (développent-ils une « professionnalité émergente » ?) et sur les élèves (font-ils des apprentissages pertinents pour leur devenir ?).

Enfin, un dernier palier reste encore à franchir pour élargir le champ de l'évaluation (*un cinquième mouvement*). En effet, comme nous l'avons déjà souligné plus haut et comme le montre le numéro 135 de la revue Administration et Éducation « Enjeux internationaux pour les professionnels de l'éducation » (AFAE, 2012)[7], les différentes régions du monde et les divers organismes internationaux ont compris l'importance de l'éducation pour le devenir économique et social du monde. Analysant les documents émis par les grandes organisations internationales entre 1990 et 2014, Akkari et Lauwerier (2015) montrent que derrière des concepts apparemment communs (qualité, bonne gouvernance, responsabilité, privatisation, benchmarking et mesure des acquis d'apprentissage), des tensions de fond existent : l'UNESCO et l'UNICEF mettent l'accent sur une vision de l'éducation, humaine et centrée sur l'enfant, tandis que la Banque mondiale et l'OCDE défendent une vision de l'éducation néo-libérale (le savoir au service du développement économique), d'où l'accent mis sur l'évaluation des compétences et les résultats de l'apprentissage. Entre temps, bien des événements ont secoué et continuent à secouer notre époque, caractérisée par l'incertitude engendrée par l'accélération des mutations de tout ordre : du climat, de la géopolitique, du monde des technologies qui transforment le monde professionnel et les comportements sociaux, des lieux de savoir et des conflits de vérité qui se multiplient. Pour faire face à ces mutations, nous n'avons pas besoin de palmarès, mais d'une saine gouvernance, respectueuse de la diversité, attentive aux besoins particuliers des

[7] Ce dossier réalisé par 22 auteurs montre l'importance d'analyser les débats éducatifs contemporains pour apporter des lumières sur le futur de l'école (Antonio Novoa), ce qui nécessite de connaître d'autres systèmes éducatifs pour mieux comprendre et agir dans le sien (Jean-Marie De Ketele). Les approches comparatives sont donc essentielles (Luisa Lombardi). Ainsi, elles permettent d'identifier les « savoirs voyageurs », c'est-à-dire ceux qui se transmettent internationalement (Roger-François Gauthier). Ou elles permettent de comprendre pourquoi les enquêtes évaluatives internationales provoquent, selon les contextes, de l'indifférence ou font référence (Jean-Claude Emin). Ce ne sont là que quelques exemples d'article de ce dossier intéressant.

différents contextes, faisant face aux incertitudes (difficultés à prédire dans un monde en accélération) et aux sentiments grandissants d'incertitude, mettant la recherche du bien commun comme finalité centrale et palliant ainsi les tentatives de domination de certains acteurs de la société (« Les ingénieurs du chaos », dit da Empoli (2019). La question centrale devient donc : Quelle évaluation, pour quelle gouvernance de l'éducation, mener au niveau international pour (1) faire face à la difficulté de prévoir les effets des changements en cours et à venir, (2) réagir de façon pertinente dans des situations extrêmes qui se multiplient, telles les guerres, les famines, les inondations, les incendies, les conflits de vérité… ? Sans doute, les instances internationales, en collaboration avec les acteurs individuels et institutionnels nationaux et locaux, devront mener, rapidement et sur le champ, des évaluations sur les effets de ces situations sur l'éducation, sur les dispositifs mis en place pour y faire face, sur leurs effets positifs et pervers… Des évaluations rétrospectives et comparatives portant sur un éventail de situations extrêmes et vécues mettront sans doute en évidence des leçons à tirer, telles l'importance jouée par certains acteurs individuels et institutionnels, la découverte des « capacités distinctives » (Dewey, 2011/1916) de ces acteurs, les mécanismes de solidarité développés… De telles évaluations devraient à leur tour inspirer les gouvernances nationales, les politiques de développement professionnel des acteurs de l'éducation, les apprentissages pertinents à développer prioritairement auprès de nos élèves.

V. Ainsi donc, l'évaluation a plus que jamais un avenir

Face à l'accélération des mutations de notre monde, l'évaluation a plus que jamais un avenir. Il ne s'agit plus guère de processus de contrôle ou d'élaboration de palmarès, mais d'un processus d'« *intelligence collective* » et d'« *émancipation* » (ne pas s'évaluer et évaluer sans, ni contre, ni pour,

mais avec les acteurs[8]) consistant à recueillir des informations pertinentes (aspect descriptif) et à les analyser (aspect réflexif) pour comprendre (donner du sens) une situation problème complexe dans toutes ses dimensions et prendre une décision autant que possible partagée pour agir. Les situations problèmes que pose l'éducation, surtout dans un monde caractérisé par l'incertitude, sont interdépendantes : une situation d'apprentissage dans une classe n'est pas sans rapport avec l'environnement et la politique de l'établissement scolaire, ni du pays, ni de l'international. Les interdépendances vont dans les deux sens. C'est pourquoi, le champ de l'évaluation a dû s'élargir et devra de plus en plus analyser les interactions entre les différents environnements et leurs politiques.

Compte tenu de la nature complexe des objets de l'évaluation, le débat n'est plus celui entre approches quantitative et qualitative, mais essentiellement et avant tout celui de la *pertinence*, tant des informations à recueillir (ne pas se tromper d'informations, qui est apte à identifier et recueillir l'information requise) que de l'analyse de celle-ci (la discussion des significations données et donc l'importance de l'intersubjectivité). Le temps est révolu où le jugement de la pertinence était le fait d'un acteur individuel (par exemple, l'enseignant dans sa classe) ou d'un acteur institutionnel (par exemple, un ministère ou une organisation internationale) ; le jugement de pertinence est ou doit être désormais assuré par une démarche d'intelligence collective, car les différents acteurs impliqués ont chacun une représentation propre de ce qui est pertinent dans leur contexte et sont amenés à partager leurs représentations pour dégager les différentes dimensions de ce qui est à évaluer.

De la même façon, si les critères de *validité* et de *fiabilité* de l'évaluation restent importants, leur vérification tient moins au développement d'un

[8] En avril 2020, la Revue Internationale d'Éducation de Sèvres, publiait un numéro spécial suite au colloque que la revue avait organisé sur « Réformer l'éducation ». Chargé d'en faire la synthèse, De Ketele (2020, p.205-233) intitulait son article « Réformer l'éducation : travailler ensemble au bien commun en développant une intelligence collective ». On relira avec intérêt les 21 articles de ce dossier, pensés et écrits avant l'apparition du Covid-19. Nous pensons particulièrement aux apports suivants : « La notion de réforme en éducation est-elle encore pertinente aujourd'hui ? » (Novoa, 2020) ; « Qu'est-ce qu'une réforme éducative réussie ? Une réflexion sociologique » (Pons, 2020) ; « Réformes éducatives : attentes et conduite du changement » (Ndoye, 2020) ; « Vers une gouvernance apprenante » (Bardi et Véran, 2020) ; et les différents articles portant un regard sur les différentes régions du monde.

arsenal sophistiqué de techniques statistiques, car les conditions requises pour le faire ne sont que rarement remplies quand l'objet à évaluer est complexe (nombreuses dimensions sont interdépendantes) ; mais elle repose à nouveau sur une analyse réflexive des représentations des acteurs impliqués. La validité étant le rapport entre ce qui est déclaré évalué et ce qui est évalué réellement, chacun d'entre eux est amené à expliciter le sens qu'il donne à ce qui est évalué ; et la fiabilité étant l'indépendance du résultat de l'évaluation par rapport à celui qui évalue et des circonstances de l'évaluation, chaque évaluateur est amené à expliciter le référentiel et les circonstances de son évaluation. En 1984, Weiss publiait « La subjectivité blanchie » et, en 2002, Gerard tenait à réhabiliter cet ouvrage en publiant un article intitulé « L'indispensable subjectivité de l'évaluation »[9]. Les représentations (les « subjectivités ») des acteurs en ce qui concerne la validité et la fiabilité font partie des données à recueillir et sont l'objet d'une réflexion partagée : l'intersubjectivité (Anadon et Guillemette, 2007) devient une démarche essentielle dans une vision épistémologique de l'évaluation axée sur la prise en compte de la complexité (Del Rey, 2013 ; Morin, 2022, 2024).

Enfin, si le but final de l'évaluation n'est pas de porter des jugements sur des acteurs individuels et institutionnels, et donc si l'évaluation n'est pas un processus de discrimination sociale, mais vise à fonder des décisions potentielles pour l'action à prendre par des acteurs différents vivant dans des contextes différents, la validité écologique devient un critère essentiel de l'évaluation et amène à parler de caractère diversifié de l'évaluation et non de discrimination. Beaucoup d'évaluations internationales se révèlent des échecs, dans le sens où les rapports ne sont même pas connus des acteurs concernés par l'action et, même lorsqu'ils sont lus, ils sont jugés ne pas avoir pris en compte les besoins et les caractéristiques de leur contexte. Ils sont qualifiés de « hors sol ». Le numéro 96 (FEI, 2024) de la Revue Internationale d'Éducation de Sèvres sur « Les données en éducation » montre qu'il est possible d'échapper à de telles critiques lorsque les acteurs de terrain sont impliqués dans le recueil et l'analyse des données. L'existence de rapports de telles initiatives situées et leur diffusion auprès des acteurs d'autres contextes auront bien plus d'impact que la publication de « rapports-palmarès ».

[9] Gerard (2002) identifie cinq types de subjectivité, non pas à éviter mais dont il faut prendre conscience : (1) le choix du type de décision que l'évaluation va devoir fonder ; (2) le choix des critères ; (3) le choix des indicateurs ; (4) le choix de la stratégie ; (5) l'examen de l'adéquation entre indicateurs et critères.

L'avenir de l'évaluation n'est pas dans la tyrannie bureaucratique de « faire du chiffre », une tyrannie devenue de plus en plus insupportable dans un monde incertain, demandant des adaptations de plus en plus rapides et adaptées aux situations. L'évaluation devient alors de plus en plus indispensable, mais comme processus d'intelligence collective (tant aux niveaux micro que méso et macro) pour répondre aux situations (souvent imprévues) et y réagir suffisamment rapidement. Le champ de l'évaluation a dû s'élargir, non seulement en termes de niveaux, mais aussi en termes de compétences et d'expertises requises.

VI. Bibliographie

Adou, J-Y. (2017). Application et limites des méthodes d'évaluation actuelles dans l'évaluation des ODD. *eVALuation Matters*, 3° trimestre.

AFAE (Association Française des Acteurs de l'Éducation) (2012). Enjeux internationaux pour les professionnels de l'Éducation. *Administration & Éducation,* (135).

Akkari, A. et Lauwerier, T. (2015). The education policies of international organizations: Specific differences and convergences. *Prospects*, *45*, (1), 141-157.

Allal, L., Cardinet, J. et Perrenoud, P. (1976). *L'évaluation formative dans un enseignement différencié*. Peter Lang.

Allal, L. et Mottier Lopez, L. (2005). Formative Assessment for Learning: A review of Publications in French. Dans *Formative Assessment – Improving in Secondary Classrooms* (p. 241-264). OECD Publications.

Anadon, M. et Guillemette, F. (2007). La recherche qualitative est-elle nécessairement inductive ? *Recherches qualitatives*, (5), 26-37.

Annan, B. et Wootton, M. (2024). Les élèves, acteurs principaux des données en éducation en Nouvelle-Zélande. *Revue Internationale d'Éducation de Sèvres*, (96), 125-138.

Bachand, C-A. et Demers, S. (2024). Pour une culture critique des données en éducation au Québec. *Revue Internationale d'Éducation de Sèvres,* (96), 93-102.

Bardi, A-M. et Véran, J-P. (2020). Vers une gouvernance apprenante. *Revue Internationale d'Éducation de Sèvres*, (83), 105-112.

Bray, M. (2024). Données manquantes : ce que l'on sait et ce qu'il faudrait savoir sur le soutien scolaire privé. *Revue Internationale d'Éducation de Sèvres,* (96), 83-92.

Butera, F., Buchs, C. et Darnon, C. (2011). *L'évaluation, une menace.* PUF.

Cardinet, J. (1986a*). Évaluation scolaire et mesure*. De Boeck.

Cardinet, J. (1986b). *Pour apprécier le travail des élèves*. De Boeck.

Da Empoli, G. (2019). *Les ingénieurs du chaos*. Lattès.

De Ketele, J-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de Pédagogie, 103*, 59-80.

De Ketele, J-M. (2013). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée, n°XV*-1, 25-37.

De Ketele, J-M. (2020). Réformer l'éducation : travailler ensemble au bien commun en développant une intelligence collective. *Revue Internationale d'Éducation de Sèvres,* (83), 205-233.

Del Rey, A. (2013). *La tyrannie de l'évaluation*. La Découverte.

Dewey, J. (2011/1916). *Démocratie et éducation,* suivi de *Expérience et éducation*. Armand Colin.

Emin, J-C. (2012). Les enquêtes internationales en France : de l'indifférence à la « référence », *Administration & Éducation,* (153).

FEI (France Éducation International) (2024). Les données en éducation. *Revue Internationale d'Éducation de Sèvres,* (96), https://journals.openedition.org/ries/15407

Felouzis, G., Maroy, C. et van Zantem, A. (2013). *Les marchés scolaires. Sociologie d'une politique publique d'éducation*. PUF.

Fernagu Oudet, S. (2012). Favoriser un environnement « capacitant » dans les organisations. Dans É. Bourgeois (dir.), *Apprendre un travail* (p.201-213). PUF.

Figari, G. (1994). *Évaluer, Quel référentiel ?* De Boeck.

Gauthier, R-F. (2012). Savoirs voyageurs, réflexion sur la transmission internationale des savoirs éducatifs. *Administration & Éducation,* (135).

Gerard, F-M. (2002). L'indispensable subjectivité de l'évaluation. *Antipodes*, (156), 26-34.

Gillet, P. (1991). L'évaluation, une question d'éthique. *Cahiers pédagogiques,* 104-106.

Gori, R., Sauret, M-J. et Abelhauser, A. (2011). *La folie évaluation. Les nouvelles fabriques de la servitude.* Les Éditions de Mille et une nuits.

Hadji, C. (1997). *L'évaluation démystifiée*. ESF éditeur.

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

Jorro, A. (2007). *Évaluation et développement professionnel*. L'Harmattan.

Jorro, A. et De Ketele, J-M. (2013). *L'engagement professionnel en éducation et formation.* De Boeck.

Jorro, A., De Ketele, J-M. et Merhan, F. (2017). *Les apprentissages professionnels accompagnés*. De Boeck.

Lheureux, M-H. et Jacob, S. (2018). La quantophrénie comme argument de résistance à l'évaluation. *Quantification et quantité*, (33).

Lombardi, L. (2012). Se comparer en éducation : pourquoi ? comment ? *Administration & Éducation,* (135).

Marcel, J-F. et Gremion, C. (2025). *Évancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation*. Cépaduès Éditions.

Merhan, F., De Ketele, J-M. et Jorro, A. (2015). *Mutations éducatives et engagement professionnel.* De Boeck.

Monteil, J-M. (2024). En France, le besoin d'une culture partagée de la production de données. *Revue Internationale d'Éducation de Sèvres,* (96), 59-70.

Morin, E. (2022). *Réveillons-nous.* Gallimard.

Morin, E. (2024). *Cheminer vers l'essentiel*. Albin Michel.

Ndoye, M. (2020). Réformes éducatives : attentes et conduite du changement. *Revue Internationale d'Éducation de Sèvres,* (83), 45-56.

Novoa, A. (2012). Lumières sur l'école. Comment interpréter les débats éducatifs contemporains dans le monde ? *Administration & Éducation*, (153).

Novoa, A. (2020). La notion de réforme est-elle encore pertinente aujourd'hui ? *Revue Internationale d'Éducation de Sèvres,* (83), 23-32.

Nunziati, G. (1988). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques, 280*.

OCDE (2017), *Transformer les écoles en « organisations apprenantes »*. https://www.men.gov.ma/Fr/Documents/Pierre-Gouëdard.pdf , consulté le 20/02/2022

Pelletier, G. (dir.) (2009). *La gouvernance en éducation. Régulation et encadrement intermédiaires*. De Boeck.

Pons, X. (2020). Qu'est-ce qu'une réforme réussie ? Une réflexion sociologique. *Revue Internationale d'Éducation de Sèvres,* (83), 33-44.

Raveaud, M. (2024). L'inspection en Angleterre : les méthodes de l'Ofsted remises en cause. *Revue Internationale d'Éducation de Sèvres,* (97), 16-19.

Reboul, O. (1992). *Les valeurs de l'éducation*. PUF.

Ricoeur, P. (2004*). Parcours de la reconnaissance*. Seuil.

Scallon, G. (2000). *Le portfolio ou dossier d'apprentissage*. Dossiers de l'Université Laval.

Sorokin, P.A. (1959). *Tendances et déboires de la sociologie américaine.* Aubier.

Véran, J-P. et Wagnon, S. (2024a). Les données en éducation : de la standardisation mondialisée à la co-construction. *Revue Internationale d'Éducation de Sèvres,* (96), 47-58.

Véran, J-P. et Wagnon, S. (2024b). Élaborer et diffuser des données fiables à tous les acteurs éducatifs et décisionnels en Afrique subsaharienne. Un entretien avec Abdel Rahamane Baba-Moussa. *Revue Internationale d'Éducation de Sèvres,* (96), 71-82.

Vial, M. (2009). *Se former pour évaluer. Se donner une problématique et élaborer des concepts*. De Boeck.

Weinstein, J. et Bravo, J. (2024). Un nouveau système de collecte et d'usage des données pour l'amélioration scolaire durant la crise du Covid-19 au Chili, *Revue Internationale d'Éducation de Sèvres,* (96), 125-138.

Weiss, J. (1984). *La subjectivité blanchie*. IRDP.

Younès, N., Gremion, C. et Sylvestre, E. (dir.). (2020). *L'évaluation, source de synergies*. Presses de l'ADMEE.

Chapitre 5. L'évaluation formative et ses reconfigurations conceptuelles

Lucie Mottier Lopez

I. L'évaluation formative, une impulsion à la création de l'ADMEE-Europe

L'évaluation formative a représenté un des vecteurs clé qui a impulsé la création et le développement de l'ADMEE-Europe. Comme retracé par l'article anniversaire des 30 ans de l'ADMEE-Europe d'Allal et al. (2017), l'évaluation formative était au centre d'un ensemble de travaux de chercheurs et chercheuses belges et suisses, ainsi que du CEPEC[1] de Lyon, dans la période précédant la création de l'ADMEE-Europe. C'est quelques-unes de ces mêmes personnes[2] qui ont formulé le projet de créer l'association. En particulier, cinq colloques ont thématisé l'évaluation formative à la suite du fameux *Handbook on formative and summative évaluation of student learning* de Bloom et al. (1971) dans le cadre de la pédagogie de maîtrise. Dans ces colloques francophones des années 1970-1980, l'évaluation formative y a été problématisée du point de vue des enjeux de l'individualisation (Mons, 1977), de l'enseignement différencié (Genève, 1978), des contextes et des démarches (Liège, 1979), des stratégies et de la généralisation (Arlon, 1981), des objectifs, compétences et capacités

[1] Centre d'études pédagogiques pour l'expérimentation et le conseil.

[2] Linda Allal, Jean Cardinet, Jean-Marie De Ketele, Charles Delorme, Yvan Tourneur, entre autres personnalités.

(Lyon, 1982), notamment. Puis, à l'occasion du premier colloque de l'ADMEE-Europe à Dijon en 1986, une controverse fameuse a marqué les travaux, reprise au colloque de 1987 : dans quelle mesure l'évaluation formative représente-t-elle un domaine de recherche à part entière, incluant ses propres modèles conceptuels ? Ou est-il préférable qu'elle soit un objet d'étude investi par les recherches en didactiques des disciplines scolaires ? Ce questionnement est caractéristique des processus en cours, à l'époque, d'institutionnalisation des « sciences de l'éducation », ainsi que des champs qui les constituent et de leurs disciplines contributives (Hostettler et Schnweuwly, 1998). Cette question d'ordre épistémologique traverse régulièrement les travaux sur l'évaluation en éducation (Mottier Lopez et Figari, 2012) portés par l'ADMEE et ses supports scientifiques de publication (revues et ouvrages collectifs). Par rapport à ce questionnement épistémologique, les travaux sur l'évaluation formative dont nous rendons compte dans ce chapitre se réclament d'un domaine à part entière de recherche en éducation et en formation (l'évaluation et la régulation des apprentissages), développant ses propres problématisations et modélisations conceptuelles, tout en assumant des emprunts à d'autres champs disciplinaires. Le but de ce texte est de présenter plusieurs modélisations, y compris anglophones, développées au fil des 40 dernières années.

II. Apports des travaux francophones jusque dans les années 2000 : l'évaluation formative « élargie »

La distinction établie par Bloom et al. (1971) pour les apprentissages[3] entre une évaluation formative et évaluation sommative a représenté une avancée majeure au XX^{e} siècle, offrant un dépassement à la vision traditionnelle de l'évaluation. Pensée essentiellement en termes d'examens par la docimologie, cette dernière est appréhendée à partir des caractéristiques de la mesure et à des fins de contrôle normatif. Dans le cadre de la pédagogie de maîtrise de Bloom et al., l'évaluation formative a pour but, quant à elle, de *remédier aux difficultés d'apprentissage* des élèves. Comme observé dans la publication de l'OCDE de 2005 par exemple, les propositions de Bloom et al. et leurs prolongements ont eu, en une

[3] Scriven (1967) avait proposé initialement cette distinction à propos de l'évaluation des programmes de formation (programmes, méthodes, matériel pédagogique).

trentaine d'années, une portée internationale, se diffusant sur tous les continents du monde.

Pour ce qui concerne les travaux en langue française sur l'évaluation formative, à la suite d'une demande de l'OCDE, Allal et Mottier Lopez (2005) ont analysé l'ensemble des articles publiés entre 1978 et 2002 dans la revue *Mesure et évaluation en éducation*[4], qui était à l'époque l'unique revue scientifique de l'ADMEE (Canada et Europe), ainsi que des chapitres dans les ouvrages publiés après des colloques de l'ADMEE et autres ouvrages en français ayant marqué le domaine (en tout 17 ouvrages). L'analyse des autrices montre que les travaux francophones se sont en partie distancés des propositions initiales de Bloom et al. Le tableau 1, complété par quelques commentaires, résume les points principaux d'élargissement identifiés par Allal et Mottier Lopez (2005).

Conception initiale	***Commentaires***	**Conception élargie**	***Commentaires***
Insertion de l'EF après une phase d'enseignement	*L'EF est décrochée de l'enseignement*	Intégration de l'EF dans toutes les situations d'apprentissage	*L'EF est intégrée aux processus d'enseignement/apprentissage*
Utilisation de tests formatifs	*Les tests papier-crayon sont principalement utilisés*	Utilisation de divers moyens de recueil d'information	*Les moyens peuvent être formels et informel, outillés, interactifs*
Feedback + correction -> remédiation	*Le retour d'information (feedback) produit par l'EF sert à corriger les erreurs des élèves qui n'atteignent pas les objectifs fixés pour tous*	Feedback + adaptation de l'enseignement -> régulation	*Le retour d'information sert à l'adaptation (régulation, différenciation) de l'enseignement à des fins de régulation des apprentissages de tous les élèves (et pas seulement des élèves en difficulté)*
Gestion de l'EF par l'enseignant	*L'EF est mise sous la seule responsabilité de l'enseignant*	Participation active des élèves à l'EF (autoévaluation au sens large)	*L'EF s'appuie sur des pratiques d'autoévaluation, d'évaluation mutuelle entre*

[4] En tout 105 articles publiés.

Conception initiale	***Commentaires***	**Conception élargie**	***Commentaires***
			élèves, de coévaluation avec l'enseignant
Maîtrise des (mêmes) objectifs par tous les élèves	*Les mêmes objectifs guident la séquence d'enseignement pour tous les élèves*	Différenciation de l'enseignement et, dans une certaine mesure, des objectifs	*L'EF prend en compte les caractéristiques et besoins différents des élèves ; elle débouche sur des dispositifs différenciés*
Remédiation bénéfique aux élèves qui ont été évalués	*Les bénéfices de l'EF s'adressent aux élèves qui ont été évalués*	Régulation à 2 niveaux : pour les élèves évalués, pour les futurs élèves (amélioration continue de l'enseignement)	*Si les résultats de l'EF concernent les élèves qui ont été évalués, ils peuvent également être exploités pour d'autres élèves*

Tableau 1 : Conception initiale de Bloom versus conception élargie de l'évaluation formative (EF) (adapté d'Allal et Mottier Lopez, 2005, p. 269)

Nous renvoyons les lecteurs/lectrices à l'article d'Allal et Mottier Lopez pour un développement de chaque point d'élargissement. Retenons ici qu'un point majeur des travaux francophones, à la suite d'un article de Cardinet (1977), est d'avoir remplacé le concept de *feedback* par celui de *régulation*, vu comme plus large (concernant tant l'apprentissage que l'enseignement) et plus dynamique par ses différents mouvements (rétroactifs, proactifs) et temporalités (immédiates, différées) (Allal, 1979, 1988). Cette exploitation du concept de régulation associé à l'évaluation formative élargie a été abondamment reprise dans les travaux subséquents et reste, encore aujourd'hui, une caractéristique (une force) des travaux francophones.

III. Conceptualisations renouvelées de l'évaluation formative dans les travaux anglophones dès les années 2000

Tout comme pour les travaux francophones, les chercheurs/chercheuses anglophones se sont attelés à revisiter les conceptions initiales de l'évaluation formative. Une de leurs critiques, notamment, concerne le terme « formatif » lui-même, jugé ambigu, ouvert à diverses interprétations, et se traduisant par l'idée d'une évaluation planifiée en même temps que l'enseignement et effectuée sous forme de fréquents micro-tests sommatifs (Broadfoot et al., 1999). La littérature de langue anglaise (en Grande-Bretagne, aux Etats-Unis, en Australie, entre autres) présente plusieurs propositions de reconfigurations conceptuelles, dont une des particularités est de proposer des approches plus *intégratives* entre évaluation formative et évaluation sommative.[5]

[5] La littérature anglophone utilise le terme « *sommative assessment* » c'est pourquoi nous le gardons ici, alors qu'une large littérature francophone tend aujourd'hui à lui préférer le terme « évaluation certificative » (pour des raisons expliquées notamment par De Ketele, 2010).

En anglais Auteurs à l'origine[6]	Traduction en français	Dimensions principales mises en évidence
Assessment for learning Assessment Reform Group, Black (1986), Black et Wiliam, (1998)	Évaluation pour apprendre ou Évaluation soutien d'apprentissage (Allal et Laveault, 2009)	Englobe toutes formes d'évaluations (formelles et informelles, formatives et sommatives) susceptibles d'encourager les élèves à s'engager dans des apprentissages à des fins de régulation et d'autorégulation ; fortement intégrées aux activités d'enseignement et d'apprentissage.
Authentic assessment Wiggins (1998)	Évaluation authentique	Les contenus, la structure, les caractéristiques et la validité des évaluations s'associent à des situations complexes du monde « réel », sans distinction entre évaluations formatives et certificatives.
Sustainable assessment Boud (2000)	Évaluation durable	Les évaluations, quelles que soient leurs formes et leurs fonctions, répondent aux besoins d'apprentissage dans des temporalités courtes et longues (apprentissage tout au long de la vie).
Assessment as learning Earl (2003)	Évaluation comme apprentissage	Les évaluations sont vues comme susceptibles de représenter des occasions d'apprentissage, particulièrement quand les élèves sont impliqués dans des autoévaluations, des évaluations entre pairs et des évaluations collaboratives formatives. L'évaluation devient objet d'apprentissage.
Learning-oriented assessment Carless et al., (2006)	Évaluation axée sur l'apprentissage	Des évaluations qui cherchent à réconcilier les fonctions formatives et sommatives pour que toutes orientent l'apprentissage des élèves dans des directions productives.

Tableau 2 : Propositions de reconfigurations conceptuelles de l'évaluation formative (repris de Mottier Lopez et al., 2021 et complétées)

Retenons ici plus spécialement l'approche *Assessment for Learning*. Celle-ci a en effet été largement développée dans un très grand nombre de contributions anglophones, puis dans d'autres communautés linguistiques dont la communauté francophone. La figure 1 expose les principes attachés à cette approche tels qu'ils ont été définis dans les années 2000.

[6] Les auteurs et autrices dans ce tableau sont cités par Mottier Lopez et al. (2021).

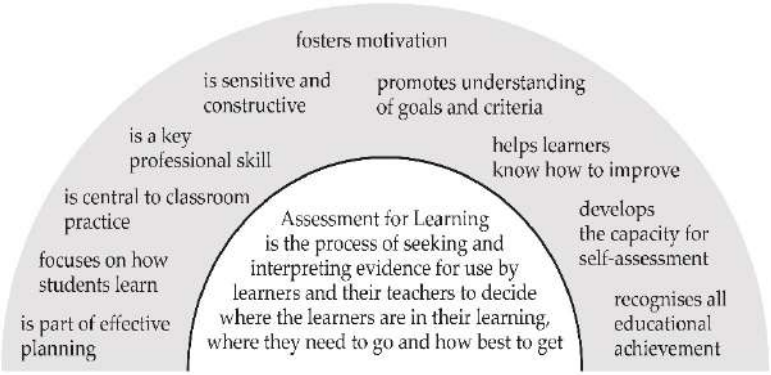


Figure 1 : Principes fondés sur la recherche pour des pratiques d'Assessment for Learning en classe (Assessment Reform Group, 2002)

Allal et Laveault (2009) résument la visée de ces principes par « des pratiques quotidiennes des élèves et des enseignants/enseignantes 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 » (p. 102). Allal et Laveault proposent de traduire en français *Assessment for Learning* par *Evaluation-soutien d'apprentissage* (ESA). Selon eux, cette approche est

> proche de celle de l'évaluation formative, qui existe en anglais et en français depuis la fin des années 1960. Elle apporte toutefois une dynamisation du concept d'évaluation formative, en soulignant l'engagement actif de l'élève dans son apprentissage, et un élargissement de champ, dans le sens que certaines démarches d'évaluation sommative (…) peuvent, selon la manière de les mener, contribuer à soutenir les apprentissages de l'élève. (p. 99)

La terminologie ESA a été reprise par des chercheurs et chercheuses qui, entre autres, ont publié dans la revue *Mesure et évaluation en éducation*, par exemple Younès et Faidit (2020), afin d'étudier des « dynamiques évaluatives à l'œuvre », ou Yerly et Monney (2023) qui interrogent la « planification flexible des démarches d'évaluation ». D'autres, dont nous faisons partie, préfèrent une traduction littérale (*évaluation pour apprendre*) visibilisant plus la filiation avec l'approche anglophone originelle. Il apparaît, en effet, que la terminologie ESA a parfois été comprise comme étant autre chose que l'*Assessment for Learning*, ce qui ne reflète pas l'intention d'Allal et Laveault (2009). Toutefois, on peut aussi penser que l'usage d'une terminologie nouvelle contribue à maintenir une dynamique

d'innovation certainement nécessaire quand, 50 ans après les travaux initiaux de Bloom et al. (1971), l'évaluation formative des apprentissages reste un objet d'étude et d'intervention.

Aujourd'hui, dans la littérature anglophone, on observe que les différentes approches présentées dans le tableau 2 tendent à se combiner et les frontières entre elles, à s'estomper. Plus généralement, nombre d'auteurs et autrices s'accordent sur l'importance des synergies à développer non seulement entre les évaluations formative et sommative/certificative (par ex., Harlen, 2007 ; Laveault et Allal, 2016 ; Mottier Lopez, 2015)[7], mais également entre les évaluations internes (produites par les enseignants/enseignantes ou équipes pédagogiques qui suivent les élèves concernés par les évaluations) et externes (produites par des acteurs/actrices du système éducatif qui n'enseignent pas aux élèves). Le terme « *Balanced Assessment* » (Stiggins, 2006) invite à penser des systèmes équilibrés et continus entre les évaluations au sein de la classe et des évaluations externes à plus grande échelle.

IV. Une évaluation continue pour apprendre durablement

On peut penser qu'une tendance actuelle est de concevoir les enjeux de l'évaluation des apprentissages des élèves (et plus largement de tout apprenant/apprenante) dans une approche intégrative plus systémique, y compris dans les travaux francophones comme en témoignent par exemple les travaux de Younès (2020) sur une « évaluation écologique » ou les travaux genevois qui étudient les pratiques d'évaluation et de régulation des apprentissages dans une perspective située (par ex., Mottier Lopez, 2006 ; Mottier Lopez, 2021a). Dans cette visée intégrative, à la suite des travaux de Linda Allal, figure marquante à l'origine de la création de l'ADMEE-Europe puis ayant œuvré pendant plus de trente ans à développer ce domaine d'étude et d'intervention, l'équipe de recherche EReD (évaluer, réguler et différencier pour apprendre dans les systèmes scolaires et de formation[8]) qui lui a succédé, met actuellement à l'épreuve

[7] Voir également les articles réunis dans le numéro 6 de La Revue LEeE « Vers des pratiques d'évaluation sommative et de notation constructives » : https://revue.leee.online/index.php/info/issue/view/13

[8] https://www.unige.ch/fapse/ered/accueil

« une évaluation continue pour apprendre durablement » (ECPA, par ex., Mottier Lopez, 2021a, 2025 ; Mottier Lopez et Girardet, 2021, 2022 ; Mottier Lopez et al., 2021), une approche qui exploite les travaux francophones et anglophones. Dans des contextes de recherche collaborative, incluant enseignants/enseignantes et élèves, cette approche invite à problématiser les pratiques d'évaluation en classe en s'appuyant sur un ensemble de fondements conceptuels présentés dans le tableau 3.

Fondements conceptuels d'une évaluation continue pour apprendre durablement	
Dialogique	Intègre dans un rapport dialogique les différents usages de l'évaluation, construisant ensemble la finalité d'une évaluation continue pour apprendre durablement
Située	N'est pas seulement centrée sur l'individu, mais considère une relation de co-constitution et de co-régulation avec les contextes éducatifs et leur ancrage socio-historique et culturel
Authentique	Répond à une activité sociale finalisée et s'appuie sur les caractéristiques des situations du monde professionnel
Continue	Est intégrée aux situations didactiques dans une succession de tâches interreliées finalisées par un but commun, permettant de révéler la façon dont les élèves apprennent au fil du temps
Durable	Articule des temporalités courtes et longues, visant à préparer les apprenants à leurs besoins professionnels futurs
Collaborative	Est partagée, coordonnée entre pairs experts et novices liés par une relation de coresponsabilité
Adaptée	Différencie les tâches évaluatives et les régulations afférentes en fonction des profils des apprenants/apprenantes
Réflexive	Intègre des processus de conscientisation critique et d'autoévaluation (au sens large) visant l'accroissement de l'autonomie et le pouvoir d'agir de l'individu et des groupes sociaux concernés
Éthique	Est formative, bienveillante, exigeante, en se fondant sur des valeurs éthiques et d'intégrité

Tableau 3 : Fondements conceptuels de l'évaluation continue pour apprendre durablement (adapté de Mottier Lopez et Girardet, 2021, p. 6)

Les cultures, les démarches, les outils, les langages, les contraintes, les obstacles sont alors autant d'objets d'étude pour appréhender, par exemple, la complexité des jugements professionnels des enseignants/enseignantes et les processus pluriels de référentialisation

qui interviennent, l'implication active des élèves pour qu'ils puissent exprimer leurs voix dans une perspective démocratique et émancipatrice (par ex., Mottier Lopez et al., 2023), les temporalités courtes et longues des trajectoires de régulation et de co-régulation vues comme potentiellement extensives (Mottier Lopez et al., à paraître) pour un apprentissage durable.

V. Perspectives critiques

Une critique adressée à la suite des propositions « opérationnelles » de Bloom et al. (1971) est le caractère prescriptif de l'évaluation formative quand celle-ci est vue comme une application de principes sans considérer la diversité des cultures, des contextes, des caractéristiques des acteurs et actrices concernés et l'hétérogénéité de leurs besoins par exemple (par ex., Allal et Mottier Lopez, 2005). Bain (1988) met en garde contre une approche instrumentale et techniciste, voire standardisée, de l'évaluation formative qui ignorerait, alors, la complexité des processus d'apprentissage et la réflexion pédagogique et didactique à mener conjointement. Cette critique du caractère prescriptif et instrumental de l'évaluation formative, virulente avant les années 2000, a conduit un ensemble d'auteurs et autrices à réaliser des recherches afin de mieux comprendre les contraintes qui pèsent sur les pratiques et leur complexité. Plusieurs obstacles ont été identifiés qui, aujourd'hui, sont encore d'actualité. Sans viser l'exhaustivité, citons ceux qui sont attachés à la distinction entre les fonctions formative et sommative/certificative de l'évaluation : soit elles sont perçues par les enseignants/enseignantes comme inconciliables, soit elles apparaissent foncièrement entremêlées dans la réalité des pratiques. Les enjeux de certification et de sélection dominent au détriment des processus formatifs (par ex., Gulikers et al., 2013 ; Romainville, 2021). Une idée répandue consiste à penser que l'évaluation formative se fait au détriment du temps dédié à l'enseignement et à l'apprentissage. Il apparaît difficile pour les enseignants/enseignantes d'identifier et d'interpréter les informations permettant d'offrir des opportunités de progression d'apprentissage (Heritage et al., 2009). L'implication des élèves dans l'évaluation ne va pas non plus de soi ; l'évaluation est susceptible de produire des effets négatifs sur leurs émotions, leur motivation, leur sentiment de compétence (Carless et Boud, 2018). Les cultures de classe et d'établissement peuvent aussi être des freins, demandant à être transformées en profondeur, avec le soutien des hiérarchies (Laveault et Allal, 2016 ; Mottier Lopez, 2023). De la

littérature actuelle, on retiendra globalement que l'implantation de l'évaluation formative est difficile et que les défis restent nombreux.

Est-ce dire que l'évaluation formative fait fausse route ? Un grand nombre d'auteurs anglophones ont mené des recherches expérimentales ou quasi-expérimentales visant à déterminer les effets de l'évaluation formative sur les apprentissages des élèves à l'école obligatoire, déjà avant les années 2000 (par ex., Black et Wiliam, 1998) et jusqu'à récemment (par ex., Sortwell et al., 2024[9]). Ces derniers constatent : « formative assessment was found to produce trivial to large positive effects on student learning, with no negative effects identified. The magnitude of effects varied according to the type of formative assessment » (p. 1). Le degré de certitude des données disponibles ne permet cependant pas, selon les auteurs, de formuler des recommandations solides concernant les stratégies à adopter plus particulièrement parmi la large palette des possibles. Plusieurs méta-analyses, qui ont fait grand bruit pour certaines (par ex., Hattie et Timperley, 2007), ont choisi de se centrer sur les effets, plus spécialement des *feedbacks* au cœur de l'évaluation formative et des processus de régulation (par ex., Mottier Lopez et al., 2021). Les analyses statistiques de Wisniewski et al. (2020)[10] mettent en évidence des tailles d'effet très différentes sur les apprentissages des élèves au regard, notamment, des différentes formes de feedbacks, des plus simples au plus complexes. Tout en invitant à une interprétation prudente, les auteurs concluent : « Feedback is more effective for cognitive and physical outcome measures than for motivational and behavioral criteria » (p. 12), sans nier l'existence aussi d'effets négatifs comme déjà constaté précédemment par Hattie et Timperley (2007). Autrement dit, tout feedback ne se vaut pas. Par extrapolation, toute évaluation formative ne se vaut pas.

Ces constats de la recherche sur les obstacles et les effets de l'évaluation formative, entre autres, appellent à poursuivre les recherches dans une perspective critique, y compris pour soutenir le développement des innovations dans un environnement en constante accélération, dont celles associées au numérique et à l'intelligence artificielle générative. Comme le rappellent Sortwell et al. (2024), pour soutenir la promotion de l'évaluation formative et de ses innovations dans les politiques et les pratiques, les décideurs, les enseignants/enseignantes, la société plus

[9] Les auteurs ont identifié et analysé de façon rigoureuse 13 méta-analyses entre 2011 et 2023 sur les effets de l'évaluation formative à l'école obligatoire.

[10] Cette étude synthétise les résultats de 435 études menées sur les effets du feedback sur les apprentissages, le comportement et la motivation.

généralement ont besoin de résultats de recherche solides et crédibles dans des environnements réels, que ce soit par le moyen de recherches expérimentales (pour mesurer des effets), compréhensives (pour mieux comprendre l'écologie des pratiques en contexte, par exemple), participatives et collaboratives avec les enseignants/enseignantes et élèves concernés (pour concevoir conjointement des dispositifs crédibles et accompagner les transformations des pratiques et des cultures évaluatives, entre autres).

En 2006 déjà, Stiggins arguait la transformation des écoles afin qu'elles ne soient plus des lieux où certains élèves réussissent et d'autres échouent inévitablement.

> Rather, they have become places where all students are expected to meet pre-specified and increasingly rigorous academic achievement standards. This change is driven by the accelerating technical and ethnic evolution of our society and the concomitant need for all students to become competent lifelong learners. (p. 1)

Nous partageons l'avis de l'auteur qui considère que toute évaluation qui produit un impact contre-productif sur l'apprentissage et/ou sur l'état émotionnel de l'élève ne peut pas être considérée comme une évaluation de qualité, même si celle-ci s'appuie sur des méthodes contrôlées et des données probantes. C'est l'esprit-même de l'évaluation formative et de ses reconfigurations conceptuelles contemporaines présentées dans ce texte, exigeant alors une éthique et de hautes compétences professionnelles chez le corps enseignant et tout autre responsable éducatif concerné (Laveault, 2012). La question de la formation des enseignants/enseignantes, de leur développement professionnel continu et de leur littératie en évaluation est évidemment cruciale. Depuis 40 ans, les travaux de l'ADMEE se sont employés à promouvoir et développer une telle perspective évaluative en sciences de l'éducation et de la formation.

VI. Bibliographie

Allal, L. (1979). Stratégies d'évaluation formative : conceptions psycho-pédagogiques et modalités d'application. Dans L. Allal, J. Cardinet, et P. Perrenoud (dir.), *L'évaluation formative dans un enseignement différencié* (p. 153-183). Peter Lang.

Allal, L. (1988). Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive. Dans M. Huberman (dir.), *Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise* (p. 86-126). Delachaux et Niestlé.

Allal, L. et Laveault, D. (2009). Assessment for learning : évaluation-soutien d'apprentissage. *Mesure et évaluation en éducation, 32*(2), 99-106. https://doi.org/10.7202/1024956ar

Allal, L. et Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. Dans

L'évaluation formative pour un meilleur apprentissage dans les classes secondaires (p. 265-290). OCDE-CERI publication. https://www.oecd.org/content/dam/oecd/fr/publications/reports/2005/01/formative-assessment_g1gh4c28/9789264007420-fr.pdf

Allal, L., Dauvisis, M.-C. et De Ketele, J.-M. (2017). L'ADMEE-Europe, née à Dijon en 1986 : développements et perspectives. *Evaluer. Journal international de Recherche en Education et Formation*, *3*(3), 107-137. https://doi.org/10.48782/d0fw5y51

Assessment Reform Group (2002). *Assessment for Learning: 10 Principles. Research-based principles to guide classroom practice Assessment for Learning*. file:///Users/mottierl/Downloads/Assessment-for-Learning-10-principles-1.pdf

Bain, D. (1988). L'évaluation formative fait fausse route. *Mesure et Evaluation en Education, 10*, 23-32.

Black, P. et Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Bloom, B. S., Hasting, J. T. et Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.

Broadfoot, P., Daugherty, R., Gardner, J., Gipps, C., Harlen, W., James, M. et Stobart, G. (1999*). Assessment for Learning: Beyond the black box*. Assessment Reform Group.

Cardinet, J. (1977). *Objectifs éducatifs et évaluation individualisée* (2e éd., Rapport n° R77.05). Institut Romand de Recherches et de Documentation Pédagogiques.

Carless, D. et Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment et Evaluation in Higher Education*, *43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée, XV*(1), 25-37. https://doi.org/10.3917/rfla.151.0025

Gulikers, J.T.M., Biemans, H.J.A., Wesselink, R. et van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation, 39*, 116-124.

HarLen, W. (2007). *Assessment of learning*. Sage publications ltd.

Hattie, J. et Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Heritage, M., Kim, J., Vendlinski, T. et Herman, J. (2009). From Evidence to Action: A Seamless Process in Formative Assessment? *Educational Measurement: Issues and Practice, 28*(3), 24-31. https://doi:10.1111/j.1745- 3992.2009.00151.x

Hofstetter, R. et Schnweuwly, B. (dir.) (1998). Le pari des sciences de l'éducation. *Raisons éducatives, 1*. https://www.unige.ch/fapse/editions/publications/re/catalogue/1998

Laveault, D. (2012). L'évaluation des compétences à enseigner : Au cœur de la professionnalisation du métier d'enseignant. *SPIRALES, 49*, 189-205.

Laveault, D. et Allal, L. (2016). Implementing Assessment for Learning: Theoretical and Practical Issues. Dans D. Laveault et L. Allal (dir.), *Assessment for Learning: Meeting the Challenge of Implementation* (p. 1-18). Springer. https://link.springer.com/book/10.1007/978-3-319-39211-0

Mottier Lopez, L. (2006). L'évaluation située des apprentissages : le rôle fondamental de la médiation sociale. Dans G. Figari et L. Mottier Lopez (dir.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie (20 ans de travaux autour de l'ADMEE-Europe)* (p. 82-89). L'Harmattan.

Mottier Lopez, L. (2015). *Evaluation formative et certificative des apprentissages. Enjeux pour l'enseignement.* De Boeck.

Mottier Lopez, L. (2021a). Évaluation située élargie : objectivation à partir de l'évaluation formative élargie, l'apprentissage situé et la référentialisation. Dans C. Barroso da Costa, I. Nizet et D. Leduc (dir.), *Quarante ans de mesure et d'évaluation* (p. 119-138). Presses Universitaires du Québec.

Mottier Lopez, L. (2021b). Une évaluation continue pour apprendre durablement, une évaluation à visée inclusive. *Revue suisse de pédagogie spécialisée, 4*, 9-16.

Mottier Lopez, L. (2023). Construire des cultures en faveur d'évaluations positives pour l'avenir : enjeux de la collaboration autour de la

production de textes à visée instituante. *La Revue LEeE, 8*. https://doi.org/10.48325/rleee.008.02

Mottier Lopez, L. (2025). Voix des élèves et microcultures de classe : enjeux pour une évancipation. Dans J.-F. Marcel et C. Gremion (dir.), *Evancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation* (p. 81-102). Cépaduès Editions.

Mottier Lopez, L. et Figari, G. (dir.) (2012). *Modélisations de l'évaluation en éducation. Questionnements épistémologiques.* De Boeck.

Mottier Lopez, L. et Girardet, C. (2021). Accompagnement doctoral à l'écriture et à l'évaluation scientifiques : objectivation du dispositif expérimenté autour des articles du numéro 3 de La Revue LEeE. *La Revue LEeE, 3.* https://doi.org/10.48325/rleee.003.08

Mottier Lopez, L. et Girardet, C. (2022). Une évaluation pour apprendre durablement : défis et dispositifs pour la formation universitaire des étudiants et des étudiantes en sciences de l'éducation. Dans M. A. Almeida de Oliveira (dir.), *Formação de professores e profissionalização docente: Práticas situadas* (p. 311-377). Pipa Comunicação.

Mottier Lopez, L., Girardet, C. et Naji, T. (2021). L'évaluation continue pour apprendre : enjeux de la pluralité des feedbacks entre pairs dans un cours universitaire. *Mesure et évaluation en éducation*, *44*(2), 1-34. https://doi.org/10.7202/1090461ar

Mottier Lopez, L., Girardet, C. et Nizet, I. (dir.) (2023). Démocratie et démocratisation des évaluations en éducation et en formation : quels enjeux ? *La Revue LEeE, 7.* https://revue.leee.online/index.php/info/issue/view/14

Mottier Lopez, L., Constenla Martinez, N., Dechamboux, L., Girardet, C., Schindelholz, M. et Simon, A. (à paraître, 2025). Evolution conceptuelle de l'évaluation formative dans ses rapports aux temps. Analyse critique. *Raisons Educatives.*

OCDE (2005). *L'évaluation formative pour un meilleur apprentissage dans les classes secondaires.* OCDE-CERI publication. https://www.oecd.org/content/dam/oecd/fr/publications/reports/2005/01/formative-assessment_g1gh4c28/9789264007420-fr.pdf

Romainville, M. (2021). Ce n'est pas en pesant le cochon qu'on le fait grossir ou pourquoi l'évaluation formative peine à se faire une place au soleil. Dans C. Barroso da Costa, D. Leduc et I. Nizet (dir.), *40 ans de mesure et d'évaluation* (p. 81-94). Presses de l'Université du Québec.

Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series on Evaluation*, *1*, 39-83.

Sortwell, A., Trimble, K., Ferraz, R., Geelan, D.R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E. et Xuan, Q. (2024). A Systematic Review of Meta-Analyses on the Impact of Formative Assessment on K-12 Students' Learning: Toward Sustainable Quality Education. *Sustainability*, *16*, 7826. https://doi.org/10.3390/su16177826

Stiggins, R. J. (2006). *Balanced assessment systems: Redefining excellence in assessment*. Educational Testing Service.

Yerly, G. et Monney, N. (2022). La planification flexible des démarches d'évaluation, un levier vers une évaluation pour apprendre ? *Mesure et évaluation en éducation, 45*(3), 33-60. https://doi.org/10.7202/1106853ar

Younès, N. (2020). *Vers une évaluation écologique dans l'enseignement supérieur : dispositifs et transformations en jeu* [Note de synthèse pour l'habilitation à diriger des recherches]. Nancy, Université de Lorraine. France.

Younès, N. et Faidit, C. (2020). Dynamiques évaluatives en jeu dans une recherche collaborative avec des enseignants et des enseignantes du second degré en France. *Mesure et évaluation en éducation, 43*(1), 67–94. https://doi.org/10.7202/1076966ar

Wisniewski, B., Zierer, K. et Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Front. Psychol. 10:3087*. https://doi: 10.3389/fpsyg.2019.03087

Chapitre 6. Evolution, tensions et perspectives de l'évaluation pour toutes et tous

Françoise Pasche Gossin

I. Introduction

L'école, en tant qu'institution centrale de nos sociétés démocratiques, joue un rôle déterminant dans la construction de la citoyenneté et la réduction des inégalités, et ceci dès l'entrée à l'école primaire jusque dans le choix des cursus de formation académique ou professionnelle. Et pourtant, on assiste, un peu partout dans nos sociétés démocratiques, à une offensive ultra conservatrice, prônant non seulement une réduction des budgets alloués à l'éducation et à la recherche, mais clamant avec force un retour des fondamentaux à l'école, comme si ceux-ci avaient été écartés ou abandonnés alors qu'ils font partie depuis toujours de nos systèmes scolaires. Plus inquiétant encore, si la massification de l'enseignement, amorcée au XIXe siècle à la suite de la révolution industrielle, a permis un lent accès à un niveau de qualification à un nombre grandissant d'enfants, elle n'a pas suffi à garantir au XXIe siècle une réelle démocratisation du système scolaire et de la formation. Les travaux de recherche montrent la persistance des inégalités, qui auraient même tendance à s'accentuer (Duru-Bellat, 2002 ; Felouzis, 2014). Dans un environnement marqué par l'incertitude, il convient de rester vigilant à certaines normes et pratiques d'évaluation, qui pourraient bien engendrer davantage d'exclusion, et compromettre une école citoyenne, émancipatrice et, plus largement, un

système de formation associé aux principes et valeurs de démocratie et d'égalité.

L'évaluation, de manière générale, est omniprésente dans les systèmes scolaires et de formation et, plus largement, dans le monde du travail. Dans le contexte scolaire, elle occupe une place importante : que l'on considère le temps passé par les enseignant·es à évaluer les apprentissages des élèves, ou bien les retombées que leurs pratiques d'évaluation peuvent avoir sur les apprentissages, la motivation et le rapport des élèves à l'école. Bien que perçue comme une pratique sociale incontournable, l'évaluation suscite des débats, des désaccords et des prises de positions parfois divergentes, dès qu'on évoque la manière d'accomplir cette tâche, de choisir les démarches et outils qui s'y réfèrent.

L'évaluation pour toutes et tous signifie que chaque élève a droit à une évaluation continue de ses apprentissages (aucun·e élève n'est exclu·e), qu'elle soit différenciée (adaptation aux profils et besoins des élèves) et qu'elle inclut des formes diversifiées d'évaluation (pluralité des formats et des modalités) (Pasche Gossin et al, 2024). Mise en œuvre notamment à travers des formes relevant du paradigme de l'évaluation-soutien d'apprentissage (EsA), l'évaluation pour toutes et tous constitue une richesse d'ouverture des possibles, mais elle présente aussi certaines limites, dont celles de nécessiter un investissement en temps conséquent ou de soulever des enjeux de faisabilité et d'égalité de traitement.

Dès lors, la question centrale à explorer dans ce texte est la suivante : l'évaluation pour toutes et tous est-elle un vœu pieu, irréaliste et utopique (dans le sens de trop idéal, difficilement réalisable) ou, au contraire, une nécessité dans le contexte sociétal et éducatif actuel et futur ? Pour répondre à cette question, le propos consiste à retracer, dans un premier temps, l'évolution de l'évaluation au cours des siècles et, surtout de ces dernières années, notamment sous l'impulsion de la recherche en évaluation en général, et dans le contexte des 40 ans d'existence de l'ADMEE-Europe. Dans un second temps, il est question non pas de s'accrocher à des certitudes, ce qui n'est certainement pas le chemin le plus fertile pour avancer, mais d'interroger les tensions persistantes au cœur de l'évaluation dans un contexte scolaire, eu égard aux défis/enjeux qui se dessinent pour nos sociétés du XXIe siècle. Même si la recherche dans le champ de l'évaluation a permis aujourd'hui des avancées notables, il subsiste encore sur le terrain professionnel des carcans idéologiques et des croyances qui entravent l'évolution des pratiques d'évaluation. En dernier lieu, il s'agit de suggérer quelques perspectives d'avenir de l'évaluation

pour toutes et tous dans nos systèmes scolaires, en réponse aux limites des évaluations uniformisées et sélectives historiquement dominantes.

Ce texte retrace de manière significative les différentes interrogations qui ont permis d'atteindre les connaissances actuelles de l'évaluation pour toutes et tous dans un contexte marqué par une quête croissante d'équité et de justice scolaire. Autrement dit, il s'agit d'examiner ces lignes de questionnement majeures qui ont jalonné les pratiques d'évaluation dans les systèmes scolaires, tout en mettant en lumière les questions vives et les défis actuels et futurs qui continueront à alimenter la démarche de l'évaluation pour toutes et tous dans une école dite démocratique.

II. Une lente émergence de l'évaluation

Le point de vue historique sur l'évaluation montre un glissement d'un paradigme de l'intuition pragmatique à celui d'évaluation-soutien d'apprentissage, en passant par l'évaluation-mesure (paradigme docimologique), suivie par l'évaluation centrée sur la maîtrise des objectifs d'apprentissage (De Ketele, 1993). Cela ne veut pas dire par exemple que le paradigme de la mesure ait disparu, mais qu'il ne peut suffire à rendre compte de la progressivité et de la complexité des apprentissages. Au fil des années, on est donc passé de pratiques d'évaluation informelles (avant le XIXe siècle) à des pratiques d'évaluation au service des apprentissages et en tant qu'apprentissage (XXIe siècle), en passant par des pratiques d'évaluation du savoir, à des fins de classement et de sélection précoce (XXe siècle). L'évolution des pratiques d'évaluation s'inscrit dans un double mouvement : d'une part, une volonté affichée de s'appuyer sur les apports des connaissances scientifiques et de diversifier les modalités d'évaluation pour construire une école démocratique, soucieuse de justice et d'attention à la diversité des publics ; d'autre part, une persistance, voire un renforcement des logiques sélectives et le constat que les acquis de la recherche peinent encore à transformer en profondeur les pratiques évaluatives effectives.

II.a. D'une pratique d'évaluation informelle à une évaluation uniforme et standardisée au XIXe siècle

L'école populaire en Suisse prend un essor vigoureux dès les années 1628. Elle est à l'origine « la servante de l'église », selon l'expression de François Guex (Maillefer, 1906, p. 44). On commença par y apprendre le catéchisme, puis l'écriture, plus tard seulement la lecture. En Suisse, l'instruction primaire obligatoire pour les filles et les garçons est ancrée dans la Constitution fédérale depuis le 29 mai 1874 :

« Les cantons pourvoient à *l'instruction primaire*, qui doit être *suffisante* et placée exclusivement *sous la direction de l'autorité civile*. Elle est *obligatoire et, dans les écoles publiques, gratuite*. Les écoles publiques doivent pouvoir être fréquentées par les adhérents de toutes les confessions, sans qu'ils aient à souffrir d'aucune façon dans leur liberté de conscience et de croyance. La Confédération prendra les mesures nécessaires contre les cantons qui ne satisferaient pas à ces obligations. » (art. 27 de la Constitution fédérale de 1874, mis en relief par la CDIP)[1].

A noter que le pouvoir d'appréciation de ce qu'est « un niveau d'instruction suffisant » est laissé aux cantons qui déterminent eux-mêmes le niveau d'une instruction suffisante.

Avant le XIXe siècle, l'évaluation telle que nous la connaissons aujourd'hui n'existait pas de manière formalisée, standardisée ou systématique. En Suisse, et dans la plupart des pays européens, elle était liée à des contextes très spécifiques (religieux, artisanaux, élitistes) et s'effectuait selon des critères souvent oraux, subjectifs et implicites. Comme le fait remarquer Guimard (2010), l'école n'évaluait pas au sens que nous l'entendons aujourd'hui, l'évaluation était informelle dans le sens où le maître jugeait par intuition et oralement les capacités des élèves. Cette évaluation relevait d'une logique d'observation et se faisait principalement à travers la récitation, la mémorisation et la capacité à reproduire fidèlement un texte (souvent religieux).

[1] https://www.edk.ch/fr/la-cdip/actualites/20240529

II.b. Développement des tests et des mesures à une remise en question et diversification de l'évaluation au XXe siècle

Le début du XXe siècle est marqué par la massification de l'enseignement, garantissant à des publics d'élèves longtemps exclus (classes populaires, filles, minorités) d'accéder à l'enseignement secondaire, puis supérieur. Cet élargissement de l'accès à l'école répond à une logique démocratique au sens où l'éducation permet l'exercice futur des droits de citoyenneté. Toutefois, si presque tous les élèves vont à l'école, cette dernière ne cherche pas vraiment à faire évoluer les pratiques évaluatives, lesquelles restent fortement basées sur les approches psychologiques, notamment sur les tests psychométriques (QI), sur des logiques de rendement et de performance et sur la mesure et la sélection (orientation précoce, filières inégalitaires). Les travaux sur la docimologie dans les années 1950 se sont fortement intéressés aux examens et à l'élaboration de la note. Une première négligence à cette époque, selon De Ketele (2006, p. 106), sera de se centrer exclusivement sur la fiabilité de la note au détriment de la pertinence de ce qu'il faut évaluer et apprendre pour certifier. Une deuxième négligence consistera, selon lui, à assimiler la fonction certificative à la démarche sommative, mais également à une démarche de classement. A noter que les travaux de Piéron en docimologie, dans les années 1960, ont montré non seulement l'existence de différents biais qui interfèrent dans la notation, mais également son manque d'information sur les acquis des élèves, ce qui la rend d'autant plus subjective et critiquable.

Il faudra attendre les années 1970 pour percevoir un tournant décisif du point de vue des recherches en évaluation. En effet, une nouvelle approche, l'évaluation formative, va se démarquer de l'évaluation traditionnelle conçue pour réaliser de simples contrôles, puis classer les élèves par des notes chiffrées. Jean Cardinet de l'Institut romand de recherche et de documentation pédagogique (IRDP, situé à Neuchâtel), Linda Allal et Philippe Perrenoud de l'Université de Genève vont organiser en 1978, à Genève, un colloque qui a pour titre *L'évaluation formative dans un enseignement différencié*, durant lequel Linda Allal va poser les bases de l'évaluation formative. A noter que le choix du titre même du colloque suggère déjà une orientation vers une évaluation différenciée, annonçant ainsi les prémisses de l'évaluation pour toutes et tous. Ce colloque ouvrira sur des échanges de plus en plus intensifs avec les chercheuses et chercheurs des pays européens voisins, la Belgique, la France, puis outre-Atlantique, avec le Canada, pour en arriver, en

septembre 1986, à la décision de fonder l'ADMEE-Europe, avec une double finalité toujours d'actualité :

> Premièrement, soutenir et stimuler des recherches scientifiques en éducation, notamment par des colloques annuels, par la publication d'ouvrages issus des colloques, par les contributions à la revue MEE, par les activités de chaque section nationale. Deuxièmement, rassembler et promouvoir un dialogue productif entre différents acteurs intéressés par l'évaluation dans le domaine de l'enseignement et de la formation : chercheurs, formateurs, conseillers pédagogiques, responsables scolaires, étudiants en science de l'éducation, enseignants en fonction (Allal et al., 2017, pp. 109-110).

Si ce besoin de réunir les travaux de la recherche en évaluation est porteur de progrès, dans les faits, on remarque que son développement sur le terrain se heurte, comme le relève Mottier Lopez (2015), à une forme de confusion entre évaluation formative et évaluation certificative sans note, ce qui tend à discréditer la fonction d'interaction et de régulation de l'évaluation formative. Cela aura également pour conséquence de provoquer, du côté des politiques éducatives, une forme d'alternance entre suppression et retour à la notation (Raulin, 2017).

II.c. D'une évaluation-soutien d'apprentissage à une évaluation pour toutes et tous au XXIe siècle

L'évaluation-soutien d'apprentissage (EsA) dans le monde francophone (Allal et Laveault, 2009) et l'assessment for learning dans le monde anglophone (Stobart, 2011) prennent une ampleur dès les années 2000. S'inscrivant dans la lignée d'une évaluation dont la fonction première est de soutenir et faire progresser les apprentissages des élèves, l'évaluation-soutien d'apprentissage élargit le champ des pratiques, en mettant l'accent sur l'implication et l'engagement actif de l'élève, sur l'importance des feed-back, sur l'évaluation différenciée et sur l'exercice d'un jugement professionnel.

L'évaluation pour toutes et tous est une terminologie assez récente qui fait écho à une école pour toutes et tous : elle participe d'une volonté politique de renforcer un idéal démocratique de l'école, selon lequel se concrétise le droit à chacun·e d'être scolarisé dans une classe ordinaire, le droit d'une école accessible à chacun·e. Le vocabulaire utilisé ici n'est donc pas neutre : il s'inscrit dans la continuité d'une école pour toutes et tous, portée par les politiques d'inclusion. Lorsqu'il s'agit d'évaluation, c'est une approche qui vise à garantir que chaque élève, indépendamment de ses différences

individuelles, puisse être évalué de manière équitable, juste et significative. Les principes sous-jacents sont l'équité et la justice, l'accessibilité, la différenciation, la bienveillance et la transparence. Ces principes mettent l'accent sur la possibilité d'adapter l'évaluation pour prendre en compte la diversité des élèves, d'utiliser des outils, des supports et des formats qui permettent à toutes et tous de répondre aux attentes de l'évaluation, d'offrir différentes façons de démontrer ses apprentissages (oral, écrit, projets, présentation, etc.), de soutenir l'apprentissage et de faire connaître auprès des élèves les critères d'évaluation. Si l'évaluation pour toutes et tous montre une évolution progressive de l'école vers plus d'équité, de justice sociale et d'attention à la diversité, il semble que le danger d'un essoufflement réside dans la récupération de cette notion à des fins de discours politiques, qui en réduisent la complexité en la vidant de sa substance pédagogique et émancipatrice.

Nous venons donc de voir que la terminologie de l'évaluation pour toutes et tous offre un degré d'élargissement différent et retient l'attention des chercheuses et chercheurs depuis quelques années déjà. Si les travaux de recherches sont nombreux à évoquer tant la pédagogie différenciée, que l'intégration et l'inclusion des élèves à besoins éducatifs particuliers, il semble que du côté de l'évaluation pour toutes et tous, les avancées portant sur cet objet témoignent d'un intérêt grandissant et prometteur. C'est ainsi que, du côté de l'ADMEE-Europe, elle contribue, elle aussi, à la réflexion en proposant lors du 35e Colloque de l'ADMEE-Europe du 24 au 26 janvier 2024, la thématique suivante : *L'évaluation face aux défis de la diversité et de l'inclusion : entre normes et différenciations*.

Au terme de ce bref détour historique, forcément partiel et loin d'être exhaustif, il convient de souligner que cette perspective a le mérite d'apporter des repères chronologiques - que nous pensons nécessaires - pour rendre compte, non seulement de la lente transformation de l'évaluation, mais aussi pour mieux comprendre les enjeux des réformes d'aujourd'hui en matière d'évaluation. Nous proposons, dans ce qui suit, d'examiner les tensions, les dilemmes, voire les paradoxes, eu égard aux défis/enjeux qui se dessinent pour nos sociétés du XXIe siècle.

III. L'évaluation pour toutes et tous au cœur de tensions dialectiques

Rendre l'évaluation accessible à l'ensemble des élèves, c'est aussi reconnaître qu'elle est traversée par de nombreux champs de tensions, des forces contraires ou des attentes différentes, sans qu'il y ait forcément un conflit ou un choix à faire. Dans cette section, le but consiste à s'attarder sur quelques tensions dialectiques, encore vives et difficiles à concilier, qui confinent pour certaines au paradoxe. Le propos n'est pas de prendre parti, mais d'exposer le fragile équilibre à trouver entre deux pôles habituellement séparés, ce qui complexifie la démarche d'évaluation pour toutes et tous, sans en diminuer la pertinence. Pour ce faire, nous proposons, avec l'espoir d'apaiser quelque peu ces tensions et non de les résoudre, quelques dispositifs d'évaluation pour toutes et tous, visant à transcender les paradigmes traditionnels d'évaluation.

III.a. Entre uniformisation et différenciation

Les enseignant·es se trouvent pris en étau entre deux logiques contradictoires : d'un côté, le besoin de rendre les évaluations uniformes (même format, mêmes critères, même outil, même cadre temporel et même barème) ; de l'autre côté, la nécessité de diversifier et différencier les outils d'évaluation aux besoins, parcours, rythmes et contextes personnels de chaque élève. Cette tension est d'autant plus marquée dans des contextes hétérogènes, où les enseignant·es sont confrontés à une diversité des publics d'élèves.

À première vue, le processus d'uniformisation des évaluations, qui consiste à évaluer les élèves de la même manière, peut sembler juste car on pense ainsi garantir une certaine forme de cohérence des apprentissages, d'égalité des droits, d'égalité de traitement et de comparabilité des résultats. Toutefois, l'école accueille des élèves aux profils variés, ne disposant pas des mêmes ressources et codes scolaires, d'où la nécessité de faire appel à des outils d'évaluation diversifiés et différenciés pour limiter les inégalités et viser l'équité. Ceci renvoie au questionnement porté à une école juste (Crahay, 2000 ; Dubet, 2004) et à une école en faveur de l'égalité et de l'équité (Demeuse et Baye, 2005 ; Meuret, 2008).

Si la tension entre uniformisation et différenciation est légitime, il semble que l'équilibre ne se trouve ni dans une uniformisation totale de

l'évaluation qui nie la diversité des profils d'élèves, ni dans une différenciation stricte de l'évaluation (individualisation) qui fragilise la cohérence des apprentissages et du système scolaire.

L'équilibre s'incarne dans une pluralité de formats, d'outils et de cadres temporels, pouvant se déployer en différentes mesures : des mesures flexibles, d'aménagement ou d'adaptation (Dubois, 2018). Ces mesures consistent à proposer aux élèves des « tâches évaluatives suffisamment flexibles » ou des « aménagements plus substantiels », tout en restant en cohérence avec les apprentissages évalués et les objectifs communs poursuivis (Pasche Gossin et al., 2024, p. 14). Quant à la mesure d'adaptation, elle fait écho à des évaluations plus centrées sur les processus d'acquisition dont les objectifs d'apprentissage sont modifiés en fonction du programme personnalisé de scolarisation de l'élève. Cela concerne en principe une catégorie d'élèves étiquetés en grandes difficultés, et fait l'objet, en Suisse et dans la plupart des pays européens, d'une décision de l'autorité scolaire.

III.b. Entre normalisation et émancipation

La tension est grande entre normalisation et émancipation. En effet, il est difficile de contester l'idée que les dispositifs d'évaluation peuvent être envisagés dans leur dimension à la fois normalisante et émancipatrice. Cette apparente évidence masque en réalité une tension bien réelle, difficilement conciliable.

D'un côté, les enseignant·es sont contraint·es de faire usage de dispositifs d'évaluation, dont l'intention peut être avant tout de comparer les élèves à une norme, à un niveau attendu ou à une moyenne qui préside au classement. D'un autre côté, l'évaluation à visée émancipatrice suppose que l'on se départisse de la logique normative de l'évaluation fondée sur la norme, pour viser à une évaluation qui implique l'élève comme acteur dans le processus, en l'associant à la construction des critères, à l'analyse de ses propres apprentissages et à la régulation de son parcours. Nous reprenons à notre compte la réflexion opérée par Caron (2022) autour du concept d'émancipation en tant que finalité pédagogique en contexte scolaire. Avant de déterminer ce que pourrait être un dispositif à visée émancipatrice, cet auteur rappelle le postulat bien connu « qu'il n'y aurait pas d'émancipation par le savoir seul » (p.13). Autrement dit, le rapport au savoir ne dépend pas uniquement de celles et ceux qui le transmettent. Il poursuit en relevant que cette vision sous l'influence des Lumières et de

Condorcet est remise en question, notamment par le fait que c'est bien le sujet en interaction avec le monde qui l'entoure qui construit ses connaissances. Dès lors, si l'on veut considérer un dispositif d'évaluation à visée émancipatrice, la double perspective, tant socioconstructiviste qu'interactionniste, apparaît donc la plus compatible. Plus concrètement, le concept d'émancipation induit que les enseignant·es soutiennent les élèves dans la co-construction d'un pouvoir d'action sur leurs propres acquisitions et compétences. Pour ce faire, certains dispositifs d'évaluation sont à même de rendre compte et d'analyser des processus d'apprentissage, plutôt que de révéler un écart à la norme. Une telle approche va induire de la part des enseignant·es un déplacement, afin de soutenir activement les élèves dans leur capacité à porter un regard concerné et critique sur leurs propres apprentissages, plutôt que de pointer les déficits, les manques ou les lacunes.

L'oscillation entre les deux pôles amène à nous intéresser à l'évaluation en tant qu'apprentissage proposée par Earl et Green (2020), dont la visée consiste à octroyer à l'élève une place centrale dans le processus d'évaluation. La question à poser est celle de savoir si les formes, bien que connues, comme l'autoévaluation, la coévaluation ou l'évaluation par les pairs, auraient pour finalité de prendre du pouvoir au sein même du processus d'évaluation, en permettant par l'expérience, à la fois le développement de l'esprit critique et la construction de compétences d'autorégulation.

III.c. Entre bienveillance et exigence

La bienveillance est souvent reliée à l'exigence dans les pratiques d'évaluation, se traduisant par la question suivante : comment fixer des attentes élevées en préservant le climat scolaire, la confiance et l'estime de soi des élèves ?

La bienveillance fait l'objet d'une injonction fréquente dans les textes officiels des différents systèmes scolaires en Suisse et dans les pays européens. Sa présence récurrente dans les discours des politiques éducatives suscite des débats, et donne lieu à de nombreuses publications en éducation (Hadji, 2012 ; Masson, 2019). Par exemple, l'enquête auprès d'enseignant·es proposée par Saillot (2018) sur la notion de bienveillance met en lumière plusieurs craintes paradoxales. Si pour certain·es, la bienveillance est une « sorte d'élixir pour penser les maux de l'école », pour d'autres, elle serait « une nouvelle excuse pour justifier n'importe quel

comportement d'élève » ou encore « une sorte de fourre-tout pédagogique ». Quant à l'exigence, elle est aussi l'objet d'ambiguïtés. Dans une certaine mesure, elle nuirait au bien-être des élèves lorsqu'elle incite à un climat de concurrence afin d'obtenir les meilleurs résultats (Sarremejane, 2017). Dans une autre mesure, elle serait indispensable lorsqu'elle porte sur les attentes à l'égard des élèves, notamment en termes d'acquisition de connaissances, de construction de compétences et de progression des apprentissages, en regard des potentialités de chaque élève.

Autrement dit, le lien qui unit tant la bienveillance que l'exigence est une obligation éducative qui porterait, selon Garel (2020), sur les apprentissages quand il s'agit d'exigence et sur le souci du bien-être de l'élève lorsqu'il s'agit de bienveillance. L'équilibre à trouver entre bienveillance et exigence à l'égard de l'évaluation serait de veiller à ce que chaque élève soit confronté à des contenus d'enseignement et à des situations d'évaluation suffisamment ambitieux et adaptés pour permettre à chacun·e de progresser.

Certaines grilles critériées, des feed-back réguliers, une évaluation en continu et intégrée, ainsi qu'un suivi constant des progrès des élèves sont autant d'éléments à combiner pour favoriser un apprentissage motivant, bienveillant et exigeant.

III.d. Entre accompagnement et certification

Il s'agit d'évoquer une dernière tension structurelle, située au cœur de l'évaluation et qui peut s'avérer vive. Parmi les finalités de l'évaluation, deux se distinguent : d'un côté, l'accompagnement des apprentissages pour valoriser les progrès, motiver les élèves à apprendre et, de l'autre côté, la certification des acquis, en produisant des notes ou des mentions, des moyennes et des bulletins à des fins institutionnelles.

Accompagner à des fins d'évaluation, c'est soutenir et réguler les apprentissages, fournir des feed-back, souligner les réussites, encourager les efforts, construire du sens, faire progresser. L'évaluation est dite en continu, intégrée, contextualisée, située, tournée vers les apprentissages et au service des apprentissages. L'évaluation-soutien d'apprentissage (EsA) est une expression largement utilisée dans le champ de l'éducation dont le but principal est de contribuer à l'apprentissage (Allal et Laveault, 2009). Certifier en évaluation, c'est poser un jugement d'acceptabilité (Hadji, 2012), valider un niveau, prendre des décisions d'ordre administratif.

L'évaluation est ponctuelle, utilisée pour orienter, attester, décider des suites du parcours scolaire et être au service de l'institution et de la société.

Si ces finalités et fonctions coexistent tant dans les discours officiels que dans les pratiques d'évaluation, leur articulation suppose de concevoir une évaluation plus intégrée, où les apprentissages sont construits en lien avec des critères explicites, partagés, co-élaborés, de sorte que la certification devienne le prolongement naturel d'un processus d'accompagnement, et non sa négation.

******

Appréhender ces champs de tensions, c'est déjà les reconnaître et tenter de trouver un équilibre entre des pôles habituellement séparés. Les pistes qui ont été ébauchées donnent à l'évaluation une couleur éthique et à volonté humaniste.

IV. Perspectives de l'évaluation pour toutes et tous

Au terme et à la lumière des tensions vives et persistances présentées ci-dessus, il s'agit d'esquisser quelques pistes concrètes à explorer pour orienter les perspectives de l'évaluation pour toutes et tous durant ces dix prochaines années.

Premièrement, il paraît important de poursuivre la dynamique des travaux de recherche sur l'évaluation. Le colloque de l'ADMEE-Europe reste en ce sens le lieu fort de rencontres des chercheuses et des chercheurs depuis quatre décennies. Cela est primordial dans le sens où la recherche contribue à produire de nouvelles connaissances en tentant de répondre à des questions encore sans réponse ou à lever le voile sur des zones d'ombre de l'évaluation. La recherche résout également des problèmes tels que l'amélioration des pratiques évaluatives des enseignant·es, également celles des formatrices et des formateurs, ou plus largement de l'usage qui en est fait dans le monde du travail. Elle éclaire également les politiques d'éducation et de formation puisqu'elle permet de faire reposer les décisions sur son positionnement et ses avancées, au détriment du sens commun. Comme le rappelle De Ketele (2006), il s'agit donc de continuer de développer des recherches dans les trois niveaux micro, méso et macro. Ceux-ci font sens puisqu'ils ont la particularité de se situer au niveau d'un processus d'enseignement-apprentissage (niveau de la classe), au niveau de l'établissement et sur le plan d'un système éducatif.

Deuxièmement, il s'agit de multiplier les recherches qui se situent du côté de l'évaluation-soutien d'apprentissage (EsA). L'évaluation est alors appréhendée comme étant au centre et au service des apprentissages. C'est une évaluation faisant usage de procédures continues, formelles et informelles, visant à promouvoir l'apprentissage, qui s'intègre au processus d'enseignement-apprentissage et au contexte spécifique de classe. La diversité des feed-back et l'implication de l'élève sont au cœur de l'EsA. À ce titre, il apparaît que la modélisation conceptuelle de l'évaluation continue pour apprendre (ECPA), développée par Mottier Lopez et Girardet (2024) va dans le prolongement de l'EsA. En effet, cette conception de l'évaluation continue pour apprendre englobe un ensemble de fondements conceptuels de l'évaluation (dialogique, située, authentique, continue, durable, collaborative, adaptée, réflexive et éthique) qu'il s'agit d'explorer, d'examiner, d'observer et de documenter non seulement à des fins scientifiques, mais également pour le développement des pratiques.

Troisièmement, il s'agit de continuer à réfléchir, comme le relève Figari (2012), au statut épistémologique de l'évaluation. Notamment, à l'objet de l'évaluation qui est découpé en autant de facettes que de disciplines ou à l'objet pris dans sa signification globale, complexe, interactive et située. La profusion des pratiques et des recherches en évaluation invite à poursuivre dans cette direction.

En dernier lieu, il y a tout intérêt à s'intéresser à ce qui entrave un chemin vers l'équité en évaluation. Une telle démarche n'implique pas de nier les différences réelles ou supposées entre les élèves, les élèves en réussite et les élèves en difficulté, mais bien de les mettre en perspective dans un contexte précis, en l'occurrence, le cadre évaluatif où elles se manifestent souvent sans ambiguïté. En tant qu'espace de construction des savoirs et d'acquisition de compétences, l'école est inscrite depuis quelques années, en Suisse et dans certains pays européens, dans une dynamique inclusive. Pour autant, l'institution scolaire est prise dans un formidable champ de tensions politiques entre une aspiration à l'équité et une nécessité de sélection. Peu encline à identifier les mécanismes liés aux inégalités qui la régissent, l'institution scolaire se montre divisée dans son approche inclusive et peu flexible dans sa capacité à questionner les modèles d'organisation de l'enseignement qu'elle impose. Ainsi, l'évaluation et les pratiques qui en découlent dépendent de ce cadre institutionnel, situé au cœur des politiques publiques. Prises en tenaille entre une tendance politique qui promeut une école inclusive et une autre plus masquée, qui donne l'impulsion à un habitus d'excellence et de performance promu par

l'école, les pratiques d'évaluation restent parfois hésitantes et peinent à trouver leur place.

V. Conclusion

La démarche proposée dans ce chapitre a d'abord retracé, au fil de l'histoire de l'évaluation, différents paradigmes : intuition pragmatique, évaluation-mesure, évaluation de la maîtrise des apprentissages et évaluation-soutien d'apprentissage. Puis, nous avons repéré quelques champs de tensions persistants qui traversent les pratiques évaluatives, tout en proposant des dispositifs ou modalités d'évaluation qui permettent à l'élève de restaurer ou d'instaurer sa capacité d'agir sur ses propres apprentissages. Enfin, nous avons proposé quelques pistes à explorer ces prochaines années. Elles concernent les actions à entreprendre pour tendre, dans les systèmes scolaires, vers une évaluation pour toutes et tous, en réponse aux limites des évaluations uniformisées et sélectives, historiquement dominantes.

Au terme de ce chapitre, il apparaît plus fondamental que jamais de poursuivre les études en évaluation, en continuant de chercher des réponses et en proposant des dispositifs d'évaluation variés, conformes aux principes éthiques de justice et d'équité pour toutes et tous.

VI. Bibliographie

Allal, L., Dauvisis, M-C. et De Ketele, J-M. (2017). L'ADMEE-Europe, née à Dijon en 1986 : développements et perspectives. Évaluer. *Journal international de Recherche en Education et Formation*, *3*(3), 107-137.

Allal, L. et Laveault, D. (2009). Assessment for Learning: évaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation*, *32*(2), 99–106. https://doi.org/10.7202/1024956ar

Caron, G. (2022). L'émancipation comme finalité pédagogique en contexte scolaire. *Penser l'éducation. Philosophie de l'éducation et histoire des idées pédagogiques*, 50, 9-34. https://doi.org/10.4000/pensereduc.608

Crahay, M. (2000). *L'école peut-elle être juste et efficace ? De l'égalité des chances et l'égalité des acquis.* De Boeck Université, Collection Pédagogie en mouvement.

Demeuse, M. et Baye, A. (2005). Pourquoi parler d'équité ? Dans M. Demeuse, A. Baye, M.H. Straeten, J. Nicaise et A. Matoul. *Vers une école juste et efficace. 26 contributions sur les systèmes d'enseignement et de formation*, (p.150-170). De Boeck Université, collection « Economie, Société, Région ». https://hal.science/hal-00808843v1

De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, *103*, 59-80. https://doi.org/10.3406/rfp.1993.1298

De Ketele, J.-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation*, *29*(1), 99–118. https://doi.org/10.7202/1086970ar

Dubet, F. (2004). *L'école des chances. Qu'est-ce qu'une école juste ?* Seuil.

Dubois, B. (2018). Oser l'évaluation différenciée. Cahiers pédagogiques, *542*(1), 35-36. https://doi.org/10.3917/cape.542.0035 .

Duru-Bellat, M. (2002). *Les inégalités sociales à l'école : Genèse et mythes.* PUF. https://doi.org/10.3917/puf.duru.2002.01

Earl, L.-M. et Green, J. (2020). L'évaluation comme apprentissage : la mettre réellement en œuvre. *Administration & Éducation*, *165*(1), 215-222. https://doi.org/10.3917/admed.165.0215.

Felouzis, G. (2014). *Les inégalités scolaires.* Presses Universitaires de France. https://doi.org/10.3917/puf.felou.2014.01

Figari, G. (2012). Configurations pour une épistémologie de l'évaluation. Dans L. Mottier Lopez et G. Figari (dir.), *Modélisations de l'évaluation en éducation : Questionnements épistémologiques* (p. 211-224). De Boeck Supérieur.

Garel, J.-P. (2020). Bienveillance et exigence pour un enseignement inclusif. *Revue Éducation, Santé, Sociétés*, *6*(2), 91-108. https://doi.org/10.17184/eac.3523

Guimard, P. (2010). *L'évaluation des compétences scolaires*. Presses universitaires de Rennes. https://doi.org/10.4000/books.pur.147

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

Maillefer, P. (1906). Histoire de l'éducation en Suisse. *Revue historique vaudoise*. https://doi.org/10.5169/seals-14600

Masson, J. (2019). Chapitre 2. La bienveillance au service de l'éducation. *Bienveillance et réussite scolaire* (p. 65-98). Dunod. https://shs.cairn.info/bienveillance-et-reussite-scolaire--9782100796199-page-65?lang=fr

Meuret, D. (2008). Entretien réalisé par Binder, C., Bodin, J. et Delahaye, J.-P. Les héritages qui pèsent sur l'école française. Humanisme, *283*(4), 49-57. https://doi.org/10.3917/huma.283.0049

Mottier Lopez. L. (2015). *Évaluation formative et certificative des apprentissages.* De Boeck.

Mottier Lopez, L. et Girardet, C. (2024). Evaluation continue pour apprendre (ECPA). Dans *Encyclopé-LEeE*. https://revue.leee.online/index.php/info/encyclope-leee/ecpa

Pasche Gossin, F., Simon, E. et Desboeufs, J.C. (2024). (*Se) former à l'évaluation-soutien d'apprentissage en cinq questions*. Collection pratique. Editions HEP-BEJUNE. ISBN 978-2-940741-03-8.

Raulin, D. (2017). Objets et pratiques de l'évaluation scolaire. *Revue française de pédagogie*, *200*(3), 69-79. https://doi.org/10.4000/rfp.7021

Saillot, E. (2018). Conforter une école bienveillante et exigeante : représentations, préoccupations et pratiques déclarées. *Questions Vives*, 29, 1-17. https://doi.org/10.4000/questionsvives.3280

Sarremejane, P. (2017). Le bien-être à l'école : analyse et épistémologie d'une notion. *Recherche & Education*. https://doi.org/10.4000/rechercheseducations.3530

Stobart, G. (2011). L'évaluation pour les apprentissages : d'une expérimentation locale à une politique nationale. *Revue française de pédagogie*, *174*(1), 41-48. https://doi.org/10.4000/rfp.2657

Chapitre 7. RCPE : des recherches collaboratives sur l'évaluation en éducation, d'une approche initialement marginale à une nouvelle alliance entre recherche et formation

Lucie Mottier Lopez, Céline Girardet

I. Création du réseau RCPE : recherches collaboratives sur les pratiques évaluatives

Le projet de la création du réseau RCPE[1] provient de la rencontre à l'Université de Genève de trois personnes, dont deux qui effectuaient à la même période un séjour scientifique dans l'équipe de recherche à laquelle appartenait Lucie Mottier Lopez : Joëlle Morrissette, à l'époque doctorante sous la direction de Serge Desgagné de l'Université Laval au Québec, et Vincent Carette de l'Université Libre de Bruxelles. A la suite de vifs débats lors d'une réunion confrontant les approches de recherche quantitatives et qualitatives, ces trois personnes se sont découvert une préoccupation commune, celle de mener des recherches avec les acteur·rices de terrain, considéré·es comme des partenaires à part entière de leurs projets, sur un même objet de recherche : les pratiques d'évaluation des apprentissages des élèves en contexte de classe. A cette époque, Vincent Carette et Lucie Mottier Lopez étaient membres du conseil d'administration de l'ADMEE-

[1] Ces éléments sont repris de Mottier Lopez (2018).

Europe, et le projet est alors né de créer un réseau thématique sur « les recherches collaboratives sur les pratiques évaluatives » (RCPE). Le choix de nommer au pluriel les recherches collaboratives visait à indiquer l'ouverture du réseau à *différentes approches collaboratives* en Europe et au Canada, et ainsi à éviter de les modéliser de façon unique et restreinte. Quant au terme « pratiques » de RCPE, il avait pour intention d'insister sur la dimension culturelle, contextuelle et située des recherches collaboratives qui ont notamment pour caractéristiques d'associer étroitement les chercheur·ses et les praticien·nes dans la réalisation de la recherche, tant du point de vue des pratiques de la recherche que des pratiques évaluatives.

Dans cette perspective, le projet soumis par Lucie Mottier Lopez et Vincent Carette au Conseil d'administration de l'ADMEE-Europe énonçait les principes suivants afin de caractériser les travaux scientifiques portés par ce nouveau réseau :

1) Une centration sur le niveau micro-contextuel des pratiques (p. ex. la classe ou un groupe de formation), tout en considérant les relations d'influence avec les autres niveaux des systèmes éducatifs et de formation.

2) Des liens entre évaluation, apprentissage et enseignement. Outre les emprunts aux disciplines contributives classiques, les apports de la pédagogie et de la didactique sont considérés comme des sources importantes de conceptualisation.

3) Des activités de co-production de connaissances entre chercheur·ses et enseignant·es/formateur·rices, alliées à un développement professionnel et contribuant à l'amélioration des pratiques enseignantes.

4) Une approche compréhensive à visée non prescriptive des pratiques étudiées, tout en visant à terme, à contribuer à leur amélioration (construction de cadres d'analyse et d'outils pour penser les pratiques). [2]

Le réseau RCPE a été officiellement adopté en juillet 2007 et a démarré à l'occasion du 20ème colloque de l'ADMEE-Europe à l'Université de Genève en 2008. Sa première réunion a eu lieu le 10 janvier 2008 réunissant des chercheur·ses belges, québécois·es et suisses.[3] Les colloques de 2009

[2] Principes copiés du projet soumis au Conseil d'administration de l'ADMEE-Europe rédigé par Mottier Lopez et Carette (juillet 2007).

[3] Avec la participation de René Barioni, Louise Bélair, Vincent Carette, Anne Clerc, Micheline-Joanne Durand, Daniel Martin, Lucie Mottier Lopez, Raphaël Pasquini, Anne Perréard Vité, Walther Tessaro, Catherine Van Nieuwenhoven.

(Louvain-la-Neuve) et de 2011 (Paris) ont permis d'élargir le nombre de membres et ont offert des opportunités pour débattre des caractéristiques des recherches collaboratives et de premiers projets de recherche s'y inscrivant.

Ainsi, par exemple à Louvain-la-Neuve[4], les membres se sont accordés sur le fait qu'observer les acteur·rices de terrain et les interviewer – « les faire parler » – ne suffisait pas à justifier une recherche collaborative. Les principes de Desgagné (1997, 2007 ; Desgagné et al., 2001) ainsi que de Bourassa et al. (2007) ont été présentés et débattus, notamment pour insister sur la co-construction de la problématique entre chercheur·ses et acteur·rices de terrain, ces dernier·ères devant être impliqué·es dans la construction des connaissances, dans l'analyse des données et dans la socialisation des résultats de la recherche. Certains membres du collectif ont jugé deux principes comme étant trop exigeants : le fait que la problématique devait forcément provenir du terrain. Plusieurs ont argué qu'elle pouvait aussi se co-construire à partir d'une question (au sens large) « apportée » par le ou la chercheur·se (sans quoi, la recherche collaborative ne pouvait répondre qu'à des attentes sociales du terrain et non pas aussi au regard de problématiques scientifiques) ; et le fait de devoir nécessairement impliquer les acteur·rices de terrain dans la diffusion des résultats de la recherche, une condition difficile à assurer dans certains contextes.

En janvier 2011, à Paris, le réseau a poursuivi sa réflexion et a notamment énoncé quatre axes pour orienter les travaux futurs : les considérations épistémologiques et théoriques sur ce courant de recherche en éducation, les implications et outils méthodologiques, les relations entre les recherches collaboratives et les enjeux d'accompagnement du développement professionnel, les articulations entre recherches collaboratives et autres approches de recherche. Cette même année, un drame a profondément marqué l'ADMEE-Europe, le décès de Vincent Carette.

Dès la première édition du réseau RCPE, un groupe de pilotage s'est constitué pour coordonner les projets.[5] Puis, dès 2015, au vu de la richesse des propositions de communications, la structure thématique du réseau s'est affinée en trois axes coordonnés par des personnes différentes : (1)

[4] Repris de la synthèse des échanges effectuée après la réunion du réseau.

[5] Louise Bélair (Québec), Vincent Carette (Belgique), Joëlle Morrissette (Québec), Lucie Mottier Lopez (Suisse) et Walther Tessaro (Suisse).

dans les contextes de la formation initiale à l'enseignement[6], (2) des enseignant·es en exercice ou avec tout autre acteur·rice professionnel·le concerné·e[7], (3) des innovations, expérimentations et projets politiques[8]. L'expansion du réseau a progressivement intégré les travaux toulousains sur la « recherche-intervention » (Marcel, 2015 ; Broussal, 2017), avec notamment les travaux scientifiques de Lucie Aussel sur les « recherches collaboratives en évaluation » (RCE). Depuis, des rencontres et des présentations régulières ont eu lieu, soit dans le cadre de symposiums RCPE dédiés, soit dans des ateliers ou autres symposiums de l'ADMEE, ainsi que dans d'autres lieux de communication scientifique et également de formation, tel l'Institut français d'éducation à Lyon en France (e.g., Monod-Ansaldi et al., 2022). La deuxième partie de ce texte vise à donner un bref aperçu des publications du réseau RCPE ainsi que des écrits sur la recherche collaborative dans les supports de publication de l'ADMEE : quels sont les questionnements abordés, sur quels objets, exploitant quelles méthodologies principalement ?

II. Analyse des articles faisant partie des numéros spéciaux RCPE, et de ceux qui ont été publiés dans les revues de l'ADMEE

Depuis la création du réseau RCPE, quatre numéros spéciaux de revues scientifiques ont été publiés explicitement associés au réseau : le premier dans la revue *e-JIREF* (2015) sur la thématique « Recherches collaboratives et évaluation : quand chercheur·ses et praticien·nes élaborent des objets d'attention communs autour de l'évaluation » (Morrissette et Tessaro, 2015). Le deuxième dans la revue *Phronesis* sur le thème « L'évaluation de dispositifs par des recherches collaboratives : enjeux d'un champ de recherche émergent » (Aussel et Mottier Lopez, 2018). Les deux autres ont été produits à la suite des 10 ans du réseau RCPE qui, à cette occasion, avait organisé un symposium « long » au 30ème colloque de l'ADMEE-Europe à Luxembourg en 2018, suivi d'un séminaire d'écriture et d'une école d'été « Évaluer les apprentissages, les enseignements, les dispositifs de formation : penser le continuum des recherches collaboratives aux

[6] Coordonné par Catherine Van Nieuwenhoven (Belgique).

[7] Coordonné par Lucie Mottier Lopez (Suisse).

[8] Coordonné par Lucie Aussel (France).

pratiques professionnelles » à l'Université de Toulouse. Ces travaux ont débouché sur le numéro spécial publié dans *La Revue LEeE* (2019) autour de la question « Quels rapports entre recherches collaboratives et évaluations ? » (Mottier Lopez, 2019), puis sur le numéro spécial dans la revue *Mesure et évaluation en éducation* (*MEE*) sur le thème « Recherches collaboratives et évaluation : de l'objet au processus » (Nizet et al., 2020).

II.a. Constitution de la base de données et description

En plus des articles faisant partie de ces quatre numéros spéciaux, nous avons recherché tous les articles qui mentionnent le mot « recherche collaborative » dans les deux revues de l'ADMEE : *MEE* et *e-JIREF.* Pour la revue *MEE*, nous avons effectué, sur ERUDIT, la recherche couvrant la période de publication de 1995 à 2024. Puis, nous avons supprimé quelques doublons (traduction d'articles en anglais dans un numéro spécial de 2022), les recensions d'ouvrages, les introductions des numéros thématiques. Ensuite, à la lecture de chaque article, nous avons supprimé ceux dans lesquels le mot « recherche collaborative » apparaissait mais sans qu'il ne désigne la démarche empirique retenue ou l'objet étudié par l'article. Nous avons appliqué le même procédé pour les articles de la revue *e-JIREF* couvrant la période de publication 2017-2024.

En tout, 35 articles ont été retenus dans notre base de données : 17 ont été publiés dans *MEE*, 12 dans *e-JIREF* et 6 dans le numéro spécial RCPE, dans *La Revue LEeE*.[9] Sur les 35 articles, 19 ne font pas partie d'un des quatre numéros spéciaux RCPE, montrant que la RC est devenue une approche méthodologique à part entière, sans être nécessairement impulsée par les productions scientifiques du réseau. La majorité de ces articles ont été publiés dans *MEE* (13 sur 19).

Les appartenances institutionnelles de l'ensemble des articles de notre base de données témoignent de l'internationalité des approches collaboratives : Canada (12), France (9), Suisse (8), Belgique (5), et un article a été co-écrit entre des auteurs et autrices d'appartenances institutionnelles canadienne et française.

[9] Pour rappel, nous n'avons pas retenu les autres articles de *La Revue LEeE* qui n'est pas directement associée à l'ADMEE.

Enfin, il est intéressant de noter que nous n'avons recensé aucun article avant la création du réseau RCPE en 2008. Dans le contexte de l'ADMEE, on peut penser que le réseau a contribué à l'expansion de cette approche méthodologique qui, près de 20 ans après, représente un courant de recherche fort en sciences de l'éducation et de la formation plus généralement (e.g., Mottier Lopez, 2022), y compris en privilégiant des désignations différentes (recherches coopératives, partenariales, participatives, interventions, orientées par la conception, recherches actions), incluant des spécificités associées.

II.b. Analyse de contenu des 35 articles

Dans un premier temps, les 35 articles ont été groupés en trois grandes catégories : (1) la RC est traitée comme objet d'étude dans l'article, (2) la RC représente la démarche méthodologique de l'objet étudié dans l'article, (3) l'article examine le rapport entre la RC et l'évaluation. Dans un deuxième temps, chaque article a fait l'objet d'une analyse de contenu réunie dans un tableau de synthèse : objet de l'article, questions de recherche, contexte(s), partenaires, méthodologie et dispositifs de recherche, temporalité de la RC, résultats en lien avec la RC, remarques notamment à propos des difficultés ou obstacles. L'annexe 1 présente la liste des articles concernés que le lectorat pourra facilement consulter en ligne. Dans notre texte ci-dessous, ils sont signalés par un astérisque. Ils ne sont pas repris dans la liste finale de références bibliographiques.

III. Résultats de l'analyse des articles retenus

Sur les 35 articles retenus, 11 étudient la RC en tant qu'objet d'étude à part entière. Trois autres interrogent le rapport entre RC et évaluation. Au vu de l'orientation de ce texte sur le réseau RCPE et ses enjeux épistémologiques et méthodologiques, nous développons ci-dessous plus spécialement les résultats qui concernent ces deux premiers ensembles d'articles. Enfin, 21 articles mentionnent la RC en tant que démarche méthodologique adoptée pour interroger diverses problématiques évaluatives en éducation et en formation.

III.a. Articles dont l'objet d'étude est la RC

Quasiment tous les articles de cette catégorie interrogent les processus de développement professionnel soutenus par les RC (7 articles sur 11), les processus d'émancipation (*Broussal, 2019) ou encore de transformation de posture chez les praticien·nes pour « se connaître comme producteurs de savoirs » (*Morrissette et Diédhiou, 2015, p. 43). Autrement dit, la RC est essentiellement examinée sous l'angle de sa visée explicitement transformatrice qui est combinée à celle de produire des savoirs scientifiques (Desgagné, 1997 ; 2007 ; Desgagné et al., 2001). Dans les articles, il est notamment observé chez les participant·es une déconstruction des représentations et des croyances à propos de l'évaluation, une évolution de celles-ci en faveur des finalités régulatrices et développementales de l'évaluation (par ex., *Hache et al., 2023 ; *Colognesi et al., 2020).

Dans une visée compréhensive et conceptuelle, les autres dimensions étudiées dans les articles sont : le déroulement des RC, les processus de collaboration entre les participant·es, les dynamiques d'ajustement et d'organisation entre eux, les difficultés rencontrées par les chercheur·ses pour mener une RC. Ce faisant, plusieurs contributions visent à développer des concepts pour toujours mieux appréhender ce qui se joue dans les RC, par exemple en termes d'objets-frontière et de fonction de « brokering » (*Hache et al., 2023 ; *Nizet et al., 2019), ou encore d'actions de collaboration observées comme étant cruciales (informer, échanger, réguler, *Cartier et al., 2015). Certain·es auteur·rices ciblent sur l'étude d'une phase en particulier de la RC, notamment celle de la « co-situation » qui sert à la coélaboration d'une problématique partagée entre les praticien·nes et les chercheur·ses (*Yerly et Kappeler, 2024). D'autres examinent l'entier de la RC pour modéliser, par exemple, sa dynamique temporelle autour de différentes étapes (incitants, conditions, processus, effets dans *Van Nieuwenhoven et Colognesi, 2015).

Sur le plan méthodologique, la plupart des chercheur·ses triangulent plusieurs méthodes de recueil et d'analyse de données. Ils s'appuient principalement sur des entretiens avec les participant·es du terrain (entretiens semi-structurés, entretiens à visée compréhensive, entretiens de groupe/focus groupes, entretiens post-recherche) et sur l'analyse des réunions collectives incluant les chercheur·ses et des traces produites dans ce contexte. Ces réunions font l'objet d'enregistrements audio, de procès-verbaux, de notes de terrain. Parfois, des questionnaires sont utilisés. De façon marginale, des observations sont faites en situation de travail et de

formation. Une recherche a par exemple utilisé une instruction au sosie. En fonction de leur projet scientifique, soit les auteur·rices s'appuient sur une recherche en particulier, soit ils et elles exploitent des données de plusieurs RC objectivées.

Enfin, souvent dans les remarques conclusives (excepté *Aussel et Bedin, 2019, qui en ont fait l'objet de leur article), des obstacles et des difficultés sont mises en évidence quand il s'agit de mener une RC : entre autres, un manque de reconnaissance et de soutien institutionnels (*Aussel et Bedin, 2019 ; *Hache et al., 2023), le temps long exigé par cette approche de recherche et la difficulté, par exemple, d'adopter une posture légitime de co-chercheur·se (*Colognesi et al., 2020), appelant plus généralement un besoin de formation spécifique (*Aussel et Bedin, 2019).

Comme le montre l'annexe 1, tous les articles dans cette catégorie, à l'exception d'un seul, font partie des numéros spéciaux émanant du réseau RCPE, révélateurs des questionnements épistémologiques et méthodologiques qui ont constamment fait l'objet d'échanges quand il s'agissait de partager et de comparer les différents travaux à l'international.

III.b. Articles qui étudient le rapport entre RC et évaluation

Comme mentionné précédemment, trois articles étudient spécifiquement le rapport entre RC et évaluation. Celui de *Leroux et al. (2020) se fonde sur une recherche d'une durée de trois ans, qui porte sur l'évaluation des compétences en formation à distance dans l'enseignement supérieur. Les auteur·rices examinent l'articulation entre la RC et les démarches évaluatives entreprises en son sein. Les résultats éclairent tout particulièrement le rôle vu comme crucial de l'évaluation à des fins de régulation pour piloter au mieux une recherche collaborative sur un temps long.

Dans une visée théorique, l'article de *Mottier Lopez et Dechamboux (2019) invite, quant à lui, à modéliser l'évaluation de la recherche collaborative en tant que « méta-enquête » ou « enquête sur l'enquête » (p. 23). Pour ce faire, l'autrice et l'auteur s'appuient sur les notions de référentialisation (Figari, 1994), d'enquête et de valuation au sens de Dewey (1938/1993, 2011). Elle et il s'appuient sur des exemples de corpus tirés de plusieurs recherches également réalisées sur des temps longs (trois et quatre ans). Les auteur·rices revendiquent une cohérence

épistémologique entre la RC et l'évaluation (de la RC, pour la RC), les amenant à proposer une modélisation particulière.

Enfin, l'article de *Marcel (2019) conceptualise les interrelations entre, plus spécifiquement, la recherche-intervention, la participation et l'évaluation. L'auteur formalise un outil d'évaluation de la recherche qui mobilise explicitement plusieurs « paradigmes évaluatifs » (p. 12) : l'évaluation-mesure, à des fins de contrôle externe ; l'évaluation-valeurs, pour un primat du sens et du vécu des acteur·rices ; l'évaluation-gestion, à des fins de régulation et de pilotage. Dans cette proposition, l'évaluation de la RC s'appuie sur des propositions connues de la littérature scientifique en les spécifiant au regard des différents buts et besoins de la RC.

D'une façon générale, ces trois articles soutiennent la conception que toutes RC, notamment quand elles se réalisent sur des temps longs, nécessitent forcément des démarches évaluatives. Un des enjeux consiste alors à toujours mieux les conceptualiser et les opérationnaliser. Pourtant, comme *Aussel et Bedin (2019) le soulignent en s'intéressant aux difficultés rencontrées par les chercheur·ses, l'évaluation sur la conduite des RC est souvent perçue comme pesante, susceptible de produire des sources de tensions ou de résistances. Une conséquence, observe plus généralement Mottier Lopez (2020), est alors d'en faire un objet « ignoré » ou euphémisé, bien que l'évaluation soit constitutive des processus collaboratifs et de ce qu'ils produisent. L'évaluation continue globalement de « faire peur » (Hadji, 2012).

III.c. Articles dont la RC est la démarche méthodologique adoptée

Pour terminer, le tableau 1 synthétise brièvement les différents objets problématisés dans les 21 articles qui mentionnent la RC comme démarche méthodologique adoptée. Une certaine focale s'observe sur l'évaluation formative (ou évaluation soutien d'apprentissage) et ses différentes dimensions problématisées.

Contextes	Objets étudiés par le moyen d'une RC
Contexte d'exercice professionnel des enseignant·es (enseignement primaire et/ou secondaire)	*Évaluation formative* Apprentissage autorégulé (*Cartier et al., 2019) Rôle du feedback (*Grangeat et Lepareur, 2019) Régulations interactives (*Mottier Lopez, 2015 ; *Mottier Lopez et Dechamboux, 2019) Co-construction d'un référentiel dans la classe (*Mottier Lopez et Dechamboux, 2019) Savoirs, pratiques et conceptions (*Morrissette, 2014 ; *Morrissette et Diédhiou, 2017 ; *Morrissette et Nadeau, 2011 ; *Ruminot, 2023) Développement professionnel (*Pasquini, 2020 ; *Tessaro, 2015 ; *Younès et Faidit, 2020) *Évaluation sommative/certificative* Alignement curriculaire (*Pasquini, 2019 ; *Pasquini, 2020) Préoccupation des enseignant·es dans des évaluations sans notes (*Saillot, 2019) Epreuves externes utilisées par les enseignant·es (*Tessaro, 2015) *Autres* Conception d'une grille critériée (focus sur l'outil) (*Wiertz et al., 2020) Éthique de la différenciation en évaluation des apprentissages (*Diédhiou et al., 2022) Reconnaissance professionnelle au sein d'un établissement scolaire (*Diédhiou et al., 2024)
Cantonal (Suisse)	Transposition de recommandations institutionnelles à propos de l'évaluation des apprentissages des élèves (*Mottier Lopez et Tessaro, 2010)
Formation professionnelle initiale	Efficience et pertinence des outils d'évaluation (*Bélair et al., 2016) Conception et usage des grilles d'évaluation et mesure des effets produits (*Deschepper, 2021) Construction des jugements évaluatifs des superviseur·ses de stages (*Maes et al., 2019)
Enseignement universitaire non professionnalisant	Co-construction de microcultures de classe en évaluation (*Soubre, 2024)

Tableau 1 : Objets étudiés par le moyen d'une RC

D'une façon générale, il ressort que la très grande majorité des recherches concernées dans cette catégorie interrogent des enjeux d'évaluation des apprentissages dans le contexte scolaire et de formation des enseignant·es.

Deux articles peuvent renvoyer au même projet de recherche collaborative (*Diédhiou et al., 2022, 2024), tandis qu'un même article peut traiter de deux recherches collaboratives distinctes (*Mottier Lopez et Tessaro, 2010)[10].

Analyser les principes de fonctionnement des RC à travers les articles s'est révélé ardu. Si la RC constitue la démarche méthodologique large dans laquelle s'inscrivent les études, les résultats s'appuient majoritairement sur une sélection de données11 en lien avec les questions de recherche de chaque article. Certains articles prennent la forme d'études de cas centrées sur un seul ou un petit nombre de participant·es, certains se focalisent sur une démarche de recueil en particulier, ou encore sur une période circonscrite. Par exemple, l'étude de *Cartier et al. (2019) s'inscrit dans le projet SLAME, qui associe chercheur·ses, enseignant·es, conseiller·ères pédagogiques, directions d'école et directions de commissions scolaires. L'étude est présentée comme une étude de cas au sein de cette recherche-action collaborative et s'appuie sur un journal de bord tenu par une enseignante, des questionnaires remplis par ses élèves et des enregistrements audio et vidéo de rencontres collectives. Si la recherche dans son ensemble s'est réalisée entre janvier 2015 et décembre 2017, cette étude s'intéresse à deux phases de la recherche ayant eu lieu entre septembre et décembre 2016. Dans ce cadre, la recherche collaborative plus large dans laquelle s'inscrit l'étude relatée dans un article n'est pas toujours décrite en détail (voire pas du tout, dans certains cas), ce qui se comprend notamment parce que ces informations ne servent pas directement les buts de l'étude, et également en raison des limites de mots/signes fixées pour la publication.

Nous avons toutefois pu dégager, à partir des informations présentes dans les articles, certaines caractéristiques méthodologiques des recherches collaboratives, sur le plan des différents partenaires impliqué·es, de l'intégration dans la recherche d'espaces-temps de rencontres entre différents types d'acteur·rices, et de méthodes de recueil de données diverses.

[10] Contrairement aux articles qui traitent de la RC comme objet d'étude, pouvant mobiliser des données issues de plusieurs RC, la quasi-totalité de ces travaux s'appuie sur des données issues d'une seule RC.

[11] Les recherches collaboratives sont reconnues comme produisant une grande quantité de données.

Du point de vue des partenaires impliqués, les RC rassemblent le plus souvent un ou quelques chercheur·ses et un petit nombre d'enseignant·es, généralement entre deux et dix. Dans le cas de collectifs plus larges, les enseignant·es sont réparti·es en sous-groupes : par « chantiers » (*Bélair et al., 2016), par degrés scolaires (*Ruminot, 2023 ; *Tessaro, 2015), par disciplines (*Saillot, 2019), ou encore par établissements (*Younès et Faidit, 2020), et ce pour tout ou certaines parties du projet collaboratif. Certaines recherches associent également d'autres types d'acteur·rices : des cadres (*Saillot, 2019), des membres de directions (*Cartier et al., 2019), des conseiller·ères pédagogiques (*Cartier et al., 2019), des formateur·rices d'enseignant·es (*Deschepper, 2021 ; Wiertz et al., 2020) ou encore des étudiant·es (*Deschepper, 2021 ; Younès et Faidit, 2020).

Dans les articles étudiés, la méthodologie est décrite en fonction des questions de recherches ciblées, ce qui relègue parfois la recherche collaborative elle-même au rang de contexte. Toutes les RC mentionnées comportent néanmoins des temps de rencontres entre chercheur·ses et acteur·rices de terrain. Ces rencontres sont désignées de diverses manières selon les auteur·rices, reflétant différentes façons de les appréhender : rencontres (*Pasquini, 2019, 2020 *Younès et Faidit, 2020), rencontres collectives (*Cartier et al., 2019), réunions (*Mottier Lopez, 2015), réunions collectives (*Mottier Lopez et Dechamboux, 2019 ; *Saillot, 2019), réunions de modération sociale (*Tessaro, 2015), entretiens collectifs (*Diédhiou et al., 2022, 2024), entretiens de groupe (*Morrissette, 2014 ; *Morrissette et Diédhiou, 2017 ; *Morrissette et Nadeau, 2011), ou groupes de discussion (*Bélair et al., 2016). Dans certains cas, ces rencontres peuvent se dérouler dans le cadre de modules de formation (*Deschepper, 2021) ou d'ateliers (*Ruminot, 2023).

Les enregistrements audio et/ou vidéo de ces rencontres constituent fréquemment des données exploitées par les études. Cependant, d'autres matériaux empiriques sont mobilisés : enregistrements d'entretiens individuels (semi-structurés, biographiques, instruction au sosie, auto-confrontation), d'entretiens collectifs ou focus groupes, journaux de bords, portfolios réflexifs, enregistrements d'observations en classe, traces écrites issues de la classe (productions d'élèves, matériel pédagogique enseignant), étude documentaire, questionnaires, notes de recherche, ou encore des supports produits au sein-même de la recherche collaborative. Cette diversité illustre la richesse des traces générées par ces dispositifs. Plus largement, la triangulation entre différentes méthodes de recueil de données au sein d'une même recherche collaborative témoigne d'une préoccupation de rigueur scientifique.

L'analyse des données est dans la grande majorité des cas qualitative. Deux exceptions recourent à des méthodes quantitatives : l'analyse statistique descriptive de réponses d'élèves à des questionnaires (*Cartier et al., 2019) et l'analyse statistique de résultats d'élèves à des tests pré et post intervention (*Deschepper, 2021).

Bien que ces articles ne problématisent pas la RC en tant qu'objet, environ la moitié d'entre eux apportent des considérations méthodologiques sur la RC, notamment en termes d'apports, de limites ou de perspectives. Ces réflexions portent principalement sur le développement professionnel, et, dans des cas isolés, sur le déroulement de la recherche, le rôle de certain·es acteur·rices, l'évolution du questionnement commun.

III.d. Perspectives

A l'origine du réseau RCPE, les RC apparaissaient marginales en Europe francophone bien que plus présentes au Québec (Vinatier et Morrissette, 2015). Dix-huit ans plus tard, la recherche collaborative dans le champ éducatif est admise, avec parfois les risques d'un « tout collaboratif »,

> sans remettre foncièrement en question le rapport hiérarchique de la recherche sur la pratique, sans non plus laisser une réelle ouverture aux apports des acteurs non académiques dans la construction du projet et des savoirs qui en résultent, sans réelles investigations conjointes ni espaces de négociation de sens à propos de l'interprétation des phénomènes investigués. (Mottier Lopez, 2018, p. 131)

Les liens entre recherche et collaboration/participation sont cependant anciens puisque la RC s'inscrit dans les développements contemporains des recherches-actions (par ex., Desgagné, 1997 ; Vinatier et Morrissette, 2015). Depuis, elles ont largement pris leur essor, visant notamment, à favoriser une nouvelle alliance entre recherche et formation, tout en revendiquant leur capacité à produire des résultats de recherche valides. Comme ce texte l'a montré, les travaux scientifiques portés par l'ADMEE contribuent activement au développement d'une telle perspective de recherche. On observe toutefois une période de moindre activité du réseau RCPE entre 2020 et 2025. Ce texte marque la volonté de revitaliser le réseau, d'en réaffirmer les fondements et d'ouvrir des perspectives autour des recherches collaboratives sur les pratiques évaluatives. Des défis importants restent en effet à relever, notamment pour assumer de proposer des épistémologies alternatives convaincantes en sciences de l'éducation

et de la formation (Rinaudo et Bedin, 2022), incluant, selon nous, des développements propres quand elles engagent explicitement des enjeux évaluatifs.

IV. Bibliographie

Aussel, L. et Mottier Lopez, L. (2018). L'évaluation de dispositifs par des recherches collaboratives : enjeux d'un champ de recherche émergent. *Phronesis*, *7*(1), 1-7. https://doi.org/10.7202/1044251ar

Bourassa, M., Bélair, L.M. et Chevalier, J. (2007). Les outils de la recherche participative. *Education et francophonie, 23*(2), 1-11. https://doi.org/10.7202/1077645ar

Broussal, D. (2017). *Recherche-intervention et accompagnement du changement en éducation : vers une approche contre-culturelle de l'émancipation* [Habilitation à Diriger les Recherches, Université de Toulouse Jean-Jaurès]. HAL science ouverte. https://hal.science/tel-01536098v1

Desgagné, S. (1997). Le concept de recherche collaborative : l'idée d'un rapprochement entre chercheurs universitaires et praticiens enseignants. *Revue des sciences de l'éducation*, *XXIII*(2), 371-393. https://doi.org/10.7202/031921ar

Desgagné, S. (2007). Le défi de coproduction de savoir en recherche collaborative : autour d'une démarche de reconstruction et d'analyse de récits de pratique enseignante. Dans M. Anadon (dir.), *La recherche participative : multiples regards* (p. 89-121). Presses de l'Université du Québec.

Desgagné, S., Bednarz, N., Lebuis P., Poirier L. et Couture, C. (2001). L'approche collaborative de recherche en éducation : un rapport nouveau à établir entre recherche et formation. *Revue des sciences de l'éducation*, *27*(1), 33-64. https://doi.org/10.7202/000305ar

Dewey, J. (1993). *Logique, la théorie de l'enquête*. Presses Universitaires de France (texte original publié en 1938).

Dewey, J. (2011). *La formation des valeurs* (traduit et présenté par A. Bidet, L. Quéré, G. Truc). La découverte.

Figari, G. (1994). *Evaluer : quel référentiel ?* De Boeck.

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

Marcel, J.-F. (dir.). (2015). *La recherche-intervention par les sciences de l'éducation. Accompagner le changement.* Educagri Editions.

Monod-Ansaldi, R., Gruson, B. et Loisy, C. (dir.). (2022). *Le réseau des LéA, un instrument pour la recherche en éducation*. Presses Universitaires de Rennes.

Morrissette, J. et Tessaro, W. (2015). Apports des recherches collaboratives sur le développement professionnel des enseignants et des formateurs en matière d'évaluation. *e-JIREF*, *1*(2), 9-14.

Mottier Lopez, L. (2018). Une approche de recherche collaborative sur les pratiques enseignantes. Partir d'une compréhension conjointement construite pour appréhender le développement professionnel des participants. Dans D. Broussal, K. Bonnaud, J.-F. Marcel et P. Sahuc (dir.), *Recherche(s) et changement(s), dialogues et relations* (p. 131-151). Éditions Cépaduès.

Mottier Lopez, L. (2019). Quels rapports entre évaluations et recherches collaboratives en éducation ? Introduction. *La Revue LEeE, 1.*

Mottier Lopez, L. (2020). Penser l'évaluation des recherches participatives en termes d'enquêtes : quels rapports avec les visées heuristique et praxéologique ? *Questions Vives, 33.* https://doi.org/10.4000/questionsvives.4804

Mottier Lopez, L. (2022). Le réseau des lieux d'éducation associés à l'Institut français de l'éducation : perspectives scientifiques des recherches collaboratives. Dans R. Monod-Ansaldi, B. Gruson et C. Loisy (dir.), *Le réseau des LéA, un instrument pour la recherche en éducation* (p. 333-342). Presses Universitaires de Rennes.

Nizet, I., Sayac, N. et Younès, N. (2020). Dossier thématique. Recherches collaboratives et évaluation : de l'objet au processus. *Mesure et évaluation en éducation, 43*(1), 1–5.

Rinaudo, J.-L. et Bedin, V. (dir.). (2022). Les épistémologies alternatives en sciences de l'éducation et de la formation. *Les dossiers des sciences de l'éducation, 48.*

Vinatier, I. et Morrissette, J. (2015). Les recherches collaboratives : enjeux et perspectives. *Revue Carrefours de l'éducation, 39*, 137-170. https://shs.cairn.info/revue-carrefours-de-l-education-2015-1-page-137?lang=fr

V. Annexe 1

*Articles retenus dans notre base de données, avec le nom des auteur·rices, l'année de publication, le pays de l'appartenance institutionnelle des auteur·rices et le support de publication (tous sont accessibles en ligne).

Auteurs	Pays	Revues
Rapports entre RC et évaluation (N=3)		
Leroux et al., 2020	CA	LEeE (ns° RCPE)
Marcel, 2019	FR	LEeE (ns° RCPE)
Mottier Lopez et Dechamboux, 2019	CH	LEeE (ns° RCPE)
RC en tant qu'objet étudié (N=11)		
Aussel et Bedin, 2019	FR	LEeE (ns° RCPE)
Broussal, 2019	FR	LEeE (ns° RCPE)
Cartier et al., 2015	CA	e-JIREF (ns° RCPE)
Colognesi et al., 2020	BE	MEE (ns° RCPE)
Hache et al., 2023	FR	MEE
Morrissette et Diédhiou, 2015	CA	e-JIREF (ns° RCPE)
Nizet et al., 2019	CA et FR	LEeE (ns° RCPE)
Nizet et Leroux, 2015	CA	e-JIREF (ns° RCPE)
Sayac, 2020	FR	MEE (ns° RCPE)
Van Nieuwenhoven et Colognesi, 2015	BE	e-JIREF (ns° RCPE)
Yerly et Kappeler, 2015	CH	e-JIREF (ns° RCPE)
RC en tant que démarche méthodologique (N=21)		
Bélair et al., 2016	CA	e-JIREF
Cartier et al., 2019	CA	e-JIREF
Deschepper, 2021	BE	e-JIREF
Diédhiou et al., 2022	CA	MEE
Diédhiou et al., 2024	CA	e-JIREF
Grangeat et Lepareur, 2019	FR	e-JIREF
Maes et al., 2019	BE	MEE
Morrissette, 2014	CA	MEE
Morrissette et Diédhiou, 2017	CA	MEE
Morrissette et Nadeau, 2011	CA	MEE
Mottier Lopez, 2015	CH	MEE
Mottier Lopez et Dechamboux, 2019	CH	e-JIREF
Mottier Lopez et Tessaro, 2010	CH	MEE
Pasquini, 2019	CH	MEE
Pasquini, 2020	CH	MEE (ns° RCPE)

Auteurs	Pays	Revues
Ruminot, 2023	CA	MEE
Saillot, 2019	FR	MEE
Soubre, 2024	FR	MEE
Tessaro, 2015	CH	e-JIREF (ns° RCPE)
Wiertz et al., 2020	BE	MEE
Younès et Faidit, 2020	FR	MEE (ns° RCPE)

ns° = numéro spécial

Chapitre 8. De l'évaluation comme objet d'étude à l'évaluation comme pratique de recherche

L'essor de la recherche-intervention évaluative, défis et perspectives

Lucie Aussel, Isabelle Bertolino, Dominique Broussal, Christelle Chauffriasse, Manon Delbreil, Pauline Hédacq

I. Introduction

L'objet de ce chapitre est de présenter la démarche de recherche-intervention évaluative, telle qu'elle a été formalisée au sein de l'un des thèmes de l'Unité Mixte de Recherche Éducation Formation Travail Savoirs (Toulouse, France). Nous resituerons cette démarche spécifique dans le cadre général des recherches menées en sciences de l'éducation et de la formation portant sur les questions d'évaluation. Nous évoquerons ce faisant des travaux qui ont fait l'objet de communications lors des différents colloques de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE-Europe, en particulier

au sein du réseau RCPE[1]). Nous nous attacherons ensuite à montrer comment la recherche-intervention évaluative s'est progressivement formalisée, sans toutefois renier son appartenance à la « famille » des recherches-interventions. Nous nous centrerons ensuite sur un point qui nous paraît central, puisqu'il concerne à la fois les processus évaluatifs et les relations au sein d'une recherche-intervention. Il s'agit de la question du pouvoir. Le quatrième temps de ce texte donnera la parole à quatre docteures. Celles-ci reviendront sur leur travail de thèse pour porter un regard réflexif sur ce qu'a impliqué la mise en œuvre d'une évaluation dans le cadre d'une recherche doctorale. Cette dernière partie vise également à illustrer la démarche de RIE (recherche-intervention évaluative) par des exemples concrets.

II. Évaluation, quand tu nous tiens

Les lecteur·rice·s se souviennent sans doute de cette fable de La Fontaine, *Le lion amoureux*. Elle se conclut par ces vers : « Amour, amour, quand tu nous tiens,/On peut bien dire : "Adieu prudence !" » (La Fontaine, 1874, p. 19-20). L'évaluation a de commun avec le sujet de cette fable son caractère séduisant, passionnant même, complexe sans aucun doute, mais tout autant insaisissable, exclusif, voire aveuglant. Si bien que l'on pourrait entendre la célèbre formule de Zarka, « l'évaluation : un pouvoir supposé savoir » (Zarka, 2009), comme une mise en garde adressée aux chercheur·se·s eux-mêmes. Nous sommes, en tant que chercheur·se·s, toujours à risque d'être agis par l'évaluation, alors même que nous pensons « produire » des connaissances sur elle. C'est que l'évaluation ne se laisse pas aisément cantonner au rang d'objet. Insaisissable et labile, elle a vite fait de retourner un discours scientifique en plaidoyer, un résultat en jugement, un·e chercheur·se impartial·e en prosélyte, en évaluateur·rice malgré iel. Nous ne saurions rendre compte ici de l'ensemble des travaux conduits en sciences de l'éducation et de la formation qui ont contribué au cours des quarante ans passés à cette forme de prudence que l'objet requiert. Nous souhaitons cependant, dans les premières pages de cette

[1] Le réseau Recherches Collaboratives sur les Pratiques Évaluatives, créé en 2008, s'adresse aux chercheur·se·s et praticien·ne·s (enseignant·e·s, formateur·rice·s, conseiller·ère·s pédagogiques, etc.) qui étudient les pratiques évaluatives dans une approche de recherche collaborative – ou avec des composantes de recherche collaborative.

contribution, évoquer quelques-unes des références qui ont nourri notre propre rapport à l'évaluation.

Comme le souligne Figari, les débats qui ont agité les chercheur·se·s à la fin des années 1980 étaient structurés par des oppositions qui renvoyaient à des paradigmes évaluatifs : contrôle/évaluation ; quantitatif/qualitatif ; résultat/processus ; externe/interne ; objet/sujet ; prescription/description. Figari ajoutait que ces débats « témoignaient notamment du dynamisme de la réflexion suscitée par l'évaluation dans le domaine éducatif et constituaient déjà une légitimation du statut d'objet de recherche spécifique de cette dernière » (Figari, 2006, p.7). Il soulignait le fait qu'en 2006, l'évaluation faisait « l'objet d'une demande très forte, à la fois institutionnelle (avec les évaluations commanditées par les ministères et les organismes de veille nationaux et internationaux) et professionnelle (avec les attentes des entreprises et des organisations vis-à-vis des dispositifs de certification, d'assurance-qualité, d'audit et de contrôle de leur fonctionnement) » (id., p. 12). Quelques années plus tôt, Lecointe soulignait les risques, face à cette demande dont les chercheur·se·s étaient destinataires de privilégier un surinvestissement technique, au détriment de l'axe éthique (Lecointe, 1997). Toujours dans une perspective de vigilance, Figari mettait en lumière le fait que les débats sur la scientificité de l'évaluation confondent généralement ce qui relève de la recherche sur les effets de l'évaluation de ce qui concerne le processus d'évaluation lui-même. L'ambiguïté est compréhensible puisqu'il existe un lien fort entre les deux activités : « la conceptualisation de la première » est souvent utilisée « au profit des choix méthodologiques de la seconde » (Figari, 2006, p.17). Les enjeux de l'évaluation sont d'autant plus importants qu'ils dépassent le stade du diagnostic pour engager des perspectives de progrès. C'est ainsi que Jorro l'envisage dans son lien avec le développement professionnel (Jorro, 2007), montrant ainsi qu'elle participe à la construction de repères pour la profession, au pilotage des dispositifs de formation, à la reconnaissance et à la responsabilisation. Abernot avance le fait que :

Pour l'éducation, l'évaluation est à la fois un moyen et une fin. Un moyen, car pour avancer, il faut constamment être renseigné de ses progrès et se sentir valorisé par eux. Mais c'est aussi une fin car c'est l'évaluation qui donne à l'homme son potentiel d'évolution. Il s'agit donc de la cultiver comme compétence ultime, elle que les philosophes antiques appelaient sagesse. (Abernot, 2013, p. 196)

C'est dans la même volonté de lier l'évaluation aux processus de délibération et de décision, y compris en intégrant une perspective

démocratique parfois absente dans les logiques évaluatives, que Bedin a formalisé les démarches d'évaluation-conseil (Bedin, 2009). Ces évolutions ont conduit à clarifier les référentiels mobilisés par des acteur·rice·s ayant des statuts et des intérêts différents, sans omettre les chercheur·se·s. Abbet et Suchaut évoquent ainsi le fait que :

Le référentiel du chercheur peut tout d'abord être caractérisé par sa nature scientifique, avec une centration sur les questions méthodologiques et théoriques en référence avec le milieu de la recherche. Pour le praticien, l'enseignant par exemple, le référentiel d'interprétation est plus pragmatique et de nature pédagogique, puisque l'intérêt pour la recherche en éducation et ses résultats est principalement lié à la pratique professionnelle en termes d'apports didactiques ou dans le domaine de la gestion de la classe au sens large ; tout ce qui pourra contribuer à mieux répondre aux besoins des élèves et, plus largement, à améliorer les conditions d'enseignement. Enfin, pour le décideur, le référentiel est clairement politique, avec une situation toutefois complexe dans la mesure où l'intérêt des recherches est en partie déterminé par le contexte temporel. (Abbet et Suchaut, 2014, p. 7)

On note l'évolution progressive vers une conception de l'évaluation qui permet aux acteur·rice·s concerné·e·s de mettre en dialogue leurs référentiels évaluatifs, puis de prendre part ensemble aux processus de délibération. Se faisant collaborative, la recherche peut avoir également pour ambition de mettre au jour les puissants phénomènes évaluatifs « qui agissent de manière plus ou moins souterraine, dans toute situation sociale » (Younès et Faidit, 2020, p. 69). Les autrices montraient en conclusion que la recherche collaborative peut permettre « de penser et de voir autrement l'évaluation, d'enrichir à la fois l'individu et le collectif » (id., p. 69). Aussi paradoxal que cela puisse paraître, certain·e·s chercheur·se·s vont jusqu'à soutenir le principe d'une évaluation émancipatrice (Baroth, 2012 ; Marcel et Gremion, 2025), ce qui pourrait apparaître comme le chaînon manquant entre recherche évaluative et recherche-intervention, chaînon que nous nous proposons de préciser dans les paragraphes qui suivent.

III. Caractérisation de la recherche-intervention évaluative au sein des démarches de recherche-intervention

Comme nous l'évoquions à l'instant, la recherche-intervention évaluative (RIE) est une forme particulière de recherche-intervention (désormais RI). Il convient par conséquent de dire quelques mots de cette dernière, afin de mieux appréhender par la suite en quoi la RIE se spécifie. Inscrite en sciences de l'éducation et de la formation, la RI est une démarche de recherche qui a pour ambition de poursuivre un double objectif : produire des connaissances scientifiques tout autant qu'accompagner le changement. Cette démarche a donné lieu à un nombre important de publications portées par des chercheur·se·s toulousain·ne·s (Bedin et Broussal, 2015 ; Broussal et al., 2015 ; Lartiguet et al., 2022 ; Marcel, 2015 ; Marcel et Broussal, 2022). Elle vise donc à produire des connaissances ainsi qu'à prévoir les conditions de leur effectivité dans une perspective de « problématisation pratique »[2]. À l'interface de la science et des milieux décisionnels ou professionnels, la RI se propose de répondre à des demandes sociales et se contractualise dans le cadre d'une commande (Aussel, 2015 ; Marcel, 2012). Sa visée explicite est d'accompagner ou de conduire un ou plusieurs changement(s) dans différents milieux d'action (établissements scolaires, médico-sociaux, associations, municipalités, entreprises, etc.), au nom du paradigme anthropologique et humaniste des sciences humaines et sociales, se distinguant de la rhétorique managériale donc. Sa mise en œuvre nécessite la création d'instances garantissant des conditions délibératives de fonctionnement (Marcel, 2020, 2021). Citons le « tiers-espace socio-scientifique » (désormais TESS) qui assure toutes les médiations nécessaires entre les différentes logiques, parfois antagonistes, structurant la RI ; le comité de pilotage (CoPil)[3] qui la représente symboliquement et politiquement ; un ou plusieurs groupes de travail

[2] Cette problématisation interroge « les conditions de l'accord fondant une coordination en raison de l'action et la place que peuvent y jouer les savoirs scientifiques » (Berthelot, 1996, p. 259).

[3] Composé du commanditaire, de certains partenaires institutionnels parfois, des chercheur·se·s-intervenant·e·s, des représentant·e·s du GT notamment.

(GT)[4] constituant le volet opérationnel de la démarche. Deux grands principes l'irriguent axiologiquement, à la manière d'une charte de qualité : la participation et l'émancipation (Marcel et Broussal, 2017) qui permettent de jouer socialement sur la gamme étendue des récepteur·trice·s de ses « résultats » (commanditaires, partenaires institutionnel·les, professionnel·le·s opérateur·rice·s, bénéficiaires). Le défi démocratique s'exprime à plein avec ces deux principes qui mettent l'accent sur la dimension collective de la RI et sur sa potentialité à reconnaître les qualités de ses participant·e·s ainsi que les bénéfices qu'iels retirent de leur engagement (prise de conscience, accentuation de la réflexivité et du sens critique, déplacement de postures, développement professionnel).

Ce socle de références a été complété plus récemment par une réflexion menée sur « l'évaluation de la RI » (Marcel et Bedin, 2018). Elle est pensée comme un processus pluriel, pouvant articuler, selon les circonstances, trois modèles évaluatifs de la mesure, de la valeur et de la gestion (Bonniol et Vial, 1997). Cette approche se présente le plus souvent comme intégrée dans le processus de conduite et d'accompagnement du changement qu'elle permet ainsi de décrypter et d'améliorer, au fur et à mesure de son déroulement. La finalité de cette évaluation de la démarche est double : apporter, grâce aux « résultats » de l'évaluation, des éléments théoriques et méthodologiques dans la perspective d'affiner et de consolider la modélisation de la RI (visée heuristique) ; optimiser les dispositifs, modes et vecteurs d'intervention (visée praxéologique).

L'émergence de problématiques évaluatives (y compris souterraines, cf. Younès et Faidit, 2020) dans les RI semble donc consubstantielle d'une démarche qui : 1) vise à accompagner le changement, ce qui renvoie à des logiques de prise de décision ; 2) repose sur une mise en œuvre complexe, ce qui qui impose des régulations multiples ; 3) sollicite la coopération d'acteur·trices ayant des expériences, des statuts et des intérêts différents. Ceci étant dit, la RIE va se caractériser par la place constitutive qu'elle donne à l'évaluation, ce qui a conduit à l'adjonction du terme « évaluative » chez différents auteur·trice·s (Ariza, 2020 ; Aussel, 2019 ; Chauffriasse et Aussel, 2022). La RIE se définit comme une RI dont l'objet de la commande fait clairement référence à l'évaluation, ce qui signifie que

[4] Composé des chercheur·se·s-intervenant·e·s, des acteur·rice·s socio-professionnel·le·s principalement. En fonction des nécessités, d'autres acteur·rice·s peuvent rejoindre le GT de manière temporaire : des représentant·e·s du CoPil, des personnes ressources de la société civile, par exemple.

la préoccupation évaluative constitue un des contenus essentiels et explicite de la convention signée entre l'équipe de recherche et le commanditaire. Dans ces conditions, l'évaluation ne s'effectue pas, de manière indépendante, dans l'espace de la sphère académique. Elle se déroule « à chaud », dans « l'ici et maintenant » de la RI, en relation directe avec les ressources et contraintes des lieux d'intervention, ainsi qu'avec les acteurs et actrices du / des terrain(s) impliqué(s) dans la démarche. Au sein de cette interaction complexe des humains et des choses, dans une urgence qu'impose souvent le fonctionnement des organisations, la situation ne manque pas d'aspérités. Outre les formes de coercition structurelles, on note des pressions inhérentes à l'évaluation, dès lors que celle-ci est finalisée et qu'elle est destinée à des utilisateur·rice·s en attente de « résultats » pertinents et efficaces. Insistons sur ce point : une évaluation menée hors du TESS ne suffit donc pas à qualifier une RI de RIE. Et c'est bien dans le contexte de la RIE que nous poserons un peu plus loin la question du pouvoir. En effet, comme nous l'avons circonscrite, l'évaluation constitue une obligation et devra donc être effectivement réalisée ; elle est susceptible de générer des résistances par peur du contrôle auquel elle est souvent assimilée tandis que l'utilisation sociale de ses résultats peut soulever des interrogations éthiques, etc.

Par ailleurs, le cadrage institutionnel que constitue la commande est ce qui permet (ou non) l'identification d'une « RIE » par la formalisation des dimensions et objets de l'évaluation qui peuvent être variés et multiples, ce que renseigne le Tableau 1 présenté ci-après.

DIMENSIONS	OBJETS D'ÉVALUATION	EXEMPLES
Ontologique	Sujets individuels, collectifs	Évaluer aptitudes, engagements, modes d'apprentissage, etc. des personnes ou des groupes
Artefactuelle	Dispositifs	Évaluer une action de formation, un projet de développement éducatif, de nouvelles règles pédagogiques…
Praxique	Pratiques individuelles, collectives	Évaluer des actes, des gestes professionnels etc. à un niveau individuel ou plus global
Épistémologique	Connaissances (théoriques ou d'action)	Évaluer de manière sommative les « résultats » d'un apprentissage, d'une expérience ou des prérequis en amont…
Processuelle	Processus générés, accélérés, limités, stoppés par la démarche de recherche mise en œuvre	Évaluer la démarche de RI elle-même et les différents changements qu'elle produit

Tableau 1 : Les différentes dimensions et objets d'évaluation au coeur d'une commande évaluative

La dimension processuelle occupe une place à part dans cet ensemble, en ce sens qu'elle est plus englobante et plus transversale que les quatre autres : évaluer le processus de RI pourrait ainsi nécessiter d'évaluer également les autres dimensions, car on peut considérer qu'elles sont intégrées dans la démarche de RI elle-même. Mais, quelle que soit la dimension de la commande évaluative prise en compte, y répondre présente des difficultés pour les chercheur·se·s-intervenant·e·s qui s'y confrontent (Aussel et Bedin, 2019), comparativement à d'autres types de commandes. Nous mobiliserons deux raisons principales pour l'expliquer :

- le statut prévalent qu'occupe l'évaluation en sciences de l'éducation et de la formation. La thématique évaluative possède effectivement son épistémologie (Mottier Lopez et Figari, 2012), ses modèles d'analyse, ses méthodologies (Figari et Mottier Lopez, 2006) (la référentialisation (Figari, 1994), la docimologie, les standards…), etc. La RIE ne pourra oblitérer ce panel considérable de références et d'outils. Reste à voir ceux qui sont compatibles avec sa modélisation propre.

- les préjugés négatifs (Zarka, 2009) sur l'évaluation qui saturent l'imaginaire collectif (contraintes sociales, procéduralisation excessive, déshumanisation des environnements, abus de pouvoir, etc.). Convaincre certains milieux d'action de s'engager dans une RIE présentée comme productrice de changements bénéfiques pour eux-mêmes peut, dès lors, relever de la gageure.

IV. L'appréhension du pouvoir au sein de la RIE

Explorer les RIE à partir de la focale conceptuelle du pouvoir interroge l'éthique et la responsabilité des chercheur·se·s-intervenant·e·s, au cœur même de leur projet de conduite et/ou d'accompagnement du changement. Ce questionnement s'inscrit dans une mise en débat des pratiques, assumant une perspective critique dont la portée vise une réflexivité et des possibles ajustements de cette activité particulière de recherche, et ce par la mise au jour de zones d'ombres pour le moment inexplorées.

IV.a. Considérations liminaires sur la question du pouvoir

Circonscrivons dès à présent, à l'aide de l'analyse de Quelquejeu (2001), ce qu'il faut entendre par « pouvoir » dans cette contribution : sa méta-analyse des travaux d'Hanna Arendt nous permet de dégager les éléments éclairants pour situer notre travail. L'auteur met en lumière la rupture que Arendt initie avec la philosophie politique traditionnelle dans sa façon de définir le pouvoir et de démêler les liens qu'il entretient avec des concepts connexes (violence, force, puissance, etc.), et qui prêtent souvent à confusion. Arendt (1967) propose de penser le « pouvoir en commun » plutôt que « le pouvoir-sur ». Pour ce faire, elle dégage tout raccourci entre l'idée de domination et celle de pouvoir : son analyse permet de penser le dépassement d'une perception traditionnelle de l'obéissance de l'humain à l'humain (soumission) vers celle d'une obéissance aux lois (consentement).

> Le pouvoir correspond à l'aptitude de l'homme à agir, et à agir de façon concertée. Le pouvoir n'est jamais une propriété individuelle ; il appartient à un groupe et constitue de lui

> appartenir aussi longtemps que ce groupe n'est pas divisé. Lorsque nous déclarons que quelqu'un est "au pouvoir", nous entendons par là qu'il a reçu d'un certain nombre de personnes le pouvoir d'agir en leur nom. (Arendt, 1967, p. 144, dans Quelquejeu, 2001, p. 516)

La philosophe présente ainsi un caractère non hiérarchique (*vs* domination) et non instrumental (*vs* violence) du pouvoir. Elle ne nie pas que les concepts qui lui sont souvent assimilés peuvent-être pensés, de façon imbriquée (les distinguer ne signifie par les séparer précise Quelquejeu). Le pouvoir peut donc être envisagé comme un phénomène collectif qui surgit, non pas tant de la rivalité que de la communication (Enegrén, 1984).

Telle que définie dans la première partie, la RIE est un terreau fertile de création de pouvoir(s) au sens d'Arendt. Dans cette perspective, nous interrogeons ce pouvoir qui ne lèserait pas la liberté (Enegrén, 1984), au sein même de cette démarche. À l'aune des défis poursuivis – participation et émancipation – nous questionnons les leviers et les dérives à l'œuvre dans ce type précis de RI. Il nous intéresse tout particulièrement de travailler ce possible oxymore : recherche-intervention évaluative. En effet, alors même qu'elle a été fondée sur des principes de recherche démocratique et citoyenne, la RIE ne peut cependant pas ignorer les influences contextuelles et idéologiques qui la traversent. Quels sont donc les enjeux de pouvoir agissant au sein des RIE ?

IV.b. La RIE en tant que dispositif : espace de liberté et de contraintes

La RIE en tant « qu'ensemble résolument hétérogène » (Foucault, 1977, p. 299) peut être considérée comme un dispositif (Aussel, 2019) tout d'abord du fait de son caractère situé (Peeters et Charlier, 1999) mais aussi car elle est structurée par la poursuite d'un changement (conduit ou accompagné) et organisée à travers l'agencement de moyens[5] (assurés par la commande) pour atteindre un but déterminé (négocié et fixé aussi dans la commande). Choisir d'étudier cette démarche *via* ce prisme du dispositif

[5]Ces moyens pouvant être de natures très différentes : humain, financier ou matériel, par exemple.

renforce l'observation de la tension dialectique qui la parcourt. En effet, la RIE mêle à la fois :

- le respect et la négociation de règles et de normes établies, en ce sens elle se définit par une dimension conformiste ;
- le possible dépassement de ces règles et normes établies, favorisé par son orientation visant à résoudre une problématique en assumant une part de création dans des logiques structurelles ; en ce sens elles se définissent par une dimension subversive.

C'est pourquoi la RIE peut se penser comme un « dispositif » teinté de cette dualité (conformité/ subversion) permettant l'existence « [d'] espace[s] de liberté et de contrainte, mi-écrit, mi-vécu. » (Audran, 2007, p. 11). Ainsi, à partir de la réponse à une demande/commande, la RIE permet de penser et d'agir en commun dans des espaces potentiels (Belin, 2001) voire des contre-espaces[6] – tels que le TESS – donnant la possibilité aux acteurs·rice·s (tant chercheur·se·s-intervenant·e·s que partenaires au sens large) d'adhérer et de respecter le cadre créé tout autant que de le détourner. Cette ligne de crête ténue entre normalisation et émancipation suppose des logiques associées au pouvoir parfois complémentaires, parfois antagoniste : comment alors se concrétise cette performance d'équilibriste ? Afin de répondre à cette question, l'objectif est de révéler les enjeux relatifs au pouvoir dans le cadre de cette démarche. Ainsi, le projet – porté par ce texte – est donc d'aller voir au-delà des intentions et de lever le voile sur la complexité de ce type particulier de recherche participative, qui supposent de « faire avec, ou de faire grâce aux autres, et au risque des autres » (Stengers, 2019, p. 25 à propos de sympoïèse[7]). Pour ce faire et pour les besoins de la démonstration, les deux volets composant les RIE seront abordés séparément : la « recherche » d'un côté, « l'intervention évaluative » de l'autre, afin de recenser un certain nombre de zones d'ombre (risques, biais possibles, etc.) et de lumière (capacitation, pouvoir d'agir, etc.).

[6] Autrement appelés hétérotopie par Foucault (Foucault, 2009, p. 198) : « en général l'hétérotopie a pour règle de juxtaposer plusieurs espaces qui, normalement, seraient, devraient être incompatibles ».

[7] L'idée de sympoïèse est associée à la philosophe Donna Haraway et pourrait se résumer ainsi : « Les vivants sont tous actifs, ils font ; mais ce qu'ils font implique, présuppose ou crée des rapports les uns avec les autres. Ils sont impliqués les uns avec les autres. Ensemble, ils font des mondes. » (Stenger, 2019, p. 25).

Ombres et lumières du volet recherche de la RIE

D'un point de vue politique et institutionnel, de façon générale et donc non particulière à la RIE, l'histoire et le développement des universités françaises constituent pour une bonne part l'influence conformiste que nous souhaitons souligner ici (part d'ombre). Sans entrer dans des détails qui nécessiteraient de plus amples explications, la culture universitaire organisée en strates distribuant inégalement le pouvoir entre ses membres façonne l'identité professionnelle des chercheur·se·s qui gravitent dans un univers de compétition (hiérarchie et classement en fonction de différents grades). Ce premier élément tend à pointer l'acceptation de certaines dispositions institutionnelles qui ont été renforcées par un processus de néolibéralisation scientifique que l'on pourrait caractériser, par exemple, par une lutte exacerbée des places ; la performance à la publication (*publish or perich*) ; la recherche incessante de financements (Dupont, 2014). Ce premier élément pointe un fonctionnement académique empreint d'un pouvoir au sens traditionnel du terme c'est-à-dire composé d'autorité, de domination, de compétition, etc. Il nous semble un niveau de contexte important à prendre en considération car il est le quotidien des équipes de recherche.

D'un point de vue symbolique et langagier, de façon plus centrée sur le procès même de la recherche, nous distinguerons le « pouvoir des concepts » et le « pouvoir de la preuve » (Stengers, 2002), tous deux faisant référence à l'autorité même des discours scientifiques construit visant à prouver et à convaincre. Il s'agit dans ce cas précis de reconnaître le poids des validités dans le procès de recherche (par exemple validité scientifique, validité sociale) pouvant exercer une possible influence coercitive induite par une posture de « sachant·e ».

D'un point de vue méthodologique, dans la même orientation que le précédent point, il s'agit de souligner le pouvoir reconnu/attribué aux outils, instruments, dispositifs de la recherche qui s'inscrivent directement dans la logique de construction de la preuve que nous évoquions précédemment (ici validité empirique). En contrepoids, les finalités épistémologiques du volet recherche de la RIE constituent davantage la part de lumière de la démarche.

D'un point de vue politique, le choix de co-construire des connaissances permet la reconnaissance de savoir pluriels ce qui, d'emblée, tente d'estomper le poids de la culture académique hiérarchisante en accordant autant d'importance aux savoirs théoriques, d'action ou expérientiels (et *in fine* aux acteur·rice·s de différents statuts qui participent à la RIE). L'on

retrouve ici le choix de la participation et la visée d'émancipation. D'un point de vue symbolique et langagier, dans la continuité du point précédent, la construction de savoirs pluriels passe par des phases de co-construction d'outils (recueil de données, dispositifs à tester sur le terrain, etc.) et de co-analyse des éléments empiriques ce qui engage les chercheur·se·s-intervenant·e·s et les membres du GT en particulier à faire un pas de côté, à se déplacer pour aller à la rencontre de l'Autre : se faire comprendre, exposer ses logiques et routines professionnelles, donc déconstruire sa posture initiale. Cette façon de se rendre disponible est indispensable à l'atteinte des objectifs du volet recherche.

Ombres et lumières du volet intervention de la RIE

Le volet intervention de type évaluatif de la RIE, indissociable de celui de la « recherche » qui a été abordé plus haut, présente lui aussi des dimensions conformistes et subversives entraînant différents enjeux de pouvoir entre les acteur·rices.

D'un point de vue politique, institutionnel et symbolique, l'intention des commanditaires de la RIE est orientée vers une action efficace et utile. Cette perspective risque d'engendrer un biais de conformité avec les référentiels déjà en place, l'évaluation pouvant jouer à plein un rôle prescriptif au regard des règles et normes en vigueur dans le contexte considéré, au risque donc de muer en contrôle. Ainsi, dans ce cas, la dimension conformiste se retrouve à travers l'orientation de régulation que supporte la visée transformative d'une RIE. L'évaluation pouvant être ainsi perçue comme un instrument au service de la performance, de la rationalité et du rendement. Cette perspective impose de fait la sanction des écarts à la norme que différents auteurs (Hadji, 2014 ; Miller et Milner, 2004 ; Zarka, 2009) dénoncent comme une dérive évaluative qui développe tant la souffrance des individus qu'une sclérose du système. Cette dimension pourrait s'opérationnaliser au sein de la RIE dès la phase de contractualisation qui institue la ou les demandes d'intervention visant à résoudre une problématique professionnelle (ou sociale). Le but étant, de la part des commanditaires – décideur·se·s des institutions concernées majoritairement – de retrouver le fonctionnement optimum de « leur » organisation.

D'un point de vue technique, la spécificité évaluative de cette RI pourrait également influencer le choix des outils, instruments et dispositifs du fait de ses propres origines, logiques et fondements politique et sociétal. L'évaluation en tant que système normé de jugements est un processus très

structuré dont le danger serait l'imposition d'un système de valeurs par les individus en charge de construire l'évaluation (référentialisation).

Néanmoins, la RIE en tant que démarche revendiquant la participation et l'émancipation de toutes les personnes qu'elle enrôle (chercheur·se·s-intervenant·e·s inclu·e·s) permet de penser des contre-pieds aux éléments développés ci-après.

D'un point de vue méthodologique, les chercheur·se·s sont garant·e·s de l'engagement des acteur·rice·s dans les espaces dédiés à la RIE afin de permettre la co-construire l'évaluation. Cette perspective collaborative permet, en théorie, de contourner le risque d'imposition d'un système de valeurs pointé ci-dessous. Par ailleurs, le processus de référentialisation entendu au sens de Figari (1994, p. 48) comme une recherche méthodique de références pertinentes qui peuvent être la fois « universelles et particulières » permet dans ce cas d'assumer une recherche de sens, dans un dépassement des chemins tracés (procédures, règles, normes) et ce grâce à la constitution et la préservation d'espace-temps « protégés » par la démarche de RIE. Ainsi, ces conditions favorisent une prise de recul sur l'existant et les expériences qui ont été vécues pouvant engendrer la réflexivité des acteur·rice·s tant sur les pratiques que sur les dispositifs. Il ne s'agit plus de craindre dans ce cas une évaluation-contrôle comme mentionné plus haut mais d'espérer une évaluation-critique « donnant de l'importance à l'analyse, caractérisée comme « participative » et ayant recours à des « référentiels adaptés à la réalité » donnée à évaluer (Figari et Remaud, 2014, p. 27). Cela revient à dire qu'il ne s'agit plus de craindre une concentration du pouvoir à une seule catégorie d'acteur·ice·s (décideur·se·s), mais d'envisager un pouvoir en partage.

V. Des recherches doctorales à l'épreuve de la RIE : quatre docteures témoignent

Dans les paragraphes qui suivent, quatre docteures de l'UMR EFTS reviennent sur leur expérience de RIE : Isabelle Bertolino, Christelle Chauffriasse, Manon Delbreil et Pauline Hédacq. Ces témoignages seront structurés de la façon suivante : a) contexte de la RIE ; b) nature de la demande/commande et évocation des éventuels livrables ; c) place que l'évaluation tenait dans cette commande ; d) déroulement de la RIE, en particulier au regard de la perspective évaluative ; e) principaux résultats

obtenus ; f) regard réflexif sur la mise en œuvre d'une évaluation dans le cadre d'une RI ; g) perspectives : questions qui émergent et qui pourraient nourrir les recherches en évaluation à venir.

V.a. Témoignage d'Isabelle Bertolino

Durant cinq ans, j'ai eu l'honneur de m'inscrire dans un parcours doctoral de formation à la recherche (Bertolino, 2024). Cette ambition a nécessité de trouver un financement. C'est ainsi que débuta, en juin 2019, une convention industrielle de formation par la recherche (CIFRE), consacrée à l'évaluation du dispositif haut-garonnais de mixité sociale dans les collèges. Cette évaluation prit la forme d'une RIE commanditée par le Conseil départemental de la Haute-Garonne et qui devait être menée en partenariat avec l'ensemble des acteur·rice·s impliqué·e·s. Mon parcours s'acheva en septembre 2024 par une soutenance de thèse, expression d'un accomplissement personnel tout autant que d'un travail de recherche.

Le déroulement de la RIE passe par différentes étapes, certaines pouvant être l'occasion de voir s'exprimer de nombreuses tensions entre les participant·e·s. N'oublions pas que selon Dewey (1967) le désir de l'enquête émerge souvent de la confrontation des acteur·rice·s a un sentiment de trouble face à une situation. Ces troubles peuvent être assez importants lorsqu'il s'agit d'évaluer une politique publique expérimentale, ce qui était le cas du dispositif mixité. L'évaluation peut donner à voir le niveau plus ou moins bon d'adéquation entre les moyens engagés et les résultats obtenus au regard de la situation de départ, nous parlerons ici de démarche de référentialisation (Figari et al., 2014). Je résumerai le déroulement de cette RIE en cinq étapes :

1. Négocier la commande avec le Conseil départemental, étape cruciale ;
2. Explorer le terrain, se présenter et présenter la démarche (durant six mois) ;
3. Mettre en place des groupes de travail, un comité de pilotage et les animer ;
4. Collecter, analyser et co-interpréter l'ensemble des données en présence des participant·e·s ;

5. Restituer les résultats et produire un rapport détaillé d'évaluation, sa synthèse pour diffusion publique ainsi qu'une thèse.

Nos résultats ont été de deux ordres. Sur le plan de la recherche, nous avons pu modéliser un écosystème de différentes dimensions interagissant sur et avec la mixité sociale dans le cadre scolaire. Sur le plan de l'amélioration du dispositif évalué, cette expérience partagée entre tous les partenaires a été contributive du renforcement de leur coopération et de la pérennisation des moyens engagés. Un peu plus d'un an après la remise de nos conclusions, le Recteur de l'académie de Toulouse et le Président du Conseil départemental de la Haute-Garonne signaient la première convention mixité de France. Quant à moi, j'ai été embauchée en qualité de Cheffe de projet chargée de la poursuite de l'évaluation.

Des enseignements que je retire de cette expérience, l'un des plus importants concerne la place des chercheur·se·s-intervenant·e·s. Ces derniers exercent une forme d'expertise particulièrement exclusive (Crozier et Friedberg, 1977) puisqu'elle est également liée à un statut, celui, dans notre cas, de doctorante. Par ailleurs, ce positionnement est renforcé par son rôle dans la mise en œuvre de la RIE dont iels maîtrisent les règles organisationnelles théorisées au sein de l'UMR EFTS (Aussel, 2016 ; Broussal et al., 2015 ; Marcel, 2010). Au regard de ce constat, on comprend à quel point il est nécessaire de rendre compte des conditions d'accompagnement et de pilotage de la RIE pour éclairer la qualité éventuellement démocratique de ses conclusions. Cette dimension éthique semble essentielle en matière d'évaluation d'une politique publique. Elle se traduit ici par les choix opérés sur le plan méthodologique et par le souci d'organiser des temps de restitution qui soient de véritables moments d'échanges avec les acteur·rice·s.

En conclusion, considérer les participant·e·s non scientifiques de la RI comme aptes à co-construire des savoirs est probablement un des éléments les plus caractéristiques de cette démarche de recherche. La participation contribue, à cet égard, à fonder son éthique et elle exige du ou de la chercheuse-intervenante de bonnes capacités relationnelles. Nous sommes loin ici de l'image du ou de la chercheur·se comme un·e ermite·sse retiré·e du cours de la vie sociale. Bien au contraire, grâce à la RIE, j'ai pu plonger au cœur de préoccupations sociales fortes et accompagner le changement par une posture critique légitime et, aussi, une nécessaire posture réflexive.

V.b. Témoignage de Christelle Chauffriasse

La RIE dont il est question a été réalisée de 2018 à 2022 dans un contexte de changement d'ingénierie de formation au sein du Service Départemental d'Incendie et de Secours de la Haute-Garonne (désormais SDIS 31). L'origine du changement (Marcel, 2014) se situe au niveau politique (Ardouin, 2013) avec, en 2013, la réforme de l'Approche Par les Compétences dans la formation des sapeur·euse·s-pompier·ère·s (SP). Cette réforme a été renforcée en 2017 avec la refonte de la filière formation. Au niveau organisationnel, son impact est fort puisque désormais la conception des formations est entièrement portée par les concepteurs pour qui cette activité est nouvelle. Nous notons également l'intégration dans les parcours de formation de nouvelles modalités (numérique, en immersion, etc.). Enfin, au niveau pédagogique, les pratiques ont nécessairement évolué tant sur le plan des postures que sur le plan de la mise en œuvre des formations.

C'est dans ce contexte de changement que le SDIS 31 a sollicité l'UMR EFTS pour être accompagné dans leur ingénierie de formation. Cette demande a abouti à une commande d'évaluation de la formation des SP conduite au sein d'une RIE. Cette RIE avait donc pour objectif de réaliser une évaluation du système de formation en vue de l'améliorer (visée transformative) et de produire des connaissances sur le processus de référentialisation (visée heuristique). Inscrite dans cette double visée, l'évaluation a été pensée dans sa complexité par la mobilisation de trois modèles d'évaluation : la mesure, la gestion et les valeurs (Marcel et Bedin, 2018). Aussi, en lien avec le principe de participation de la RI, l'évaluation a été co-construite de manière collaborative (Dochy et al., 1999) avec un groupe de travail constitué de sept personnes et d'un comité de pilotage. L'évaluation a été au centre de la RI en servant d'une part le volet intervention et d'autre part le volet recherche avec un tiers-espace socio-scientifique permettant une relation d'interdépendance entre les sphères de la recherche et de l'intervention.

Dans un premier temps, durant la phase du « en-commun » de la RI (Marcel, 2020) nous avons réalisé un diagnostic de la formation du SDIS 31 à partir d'enquêtes qualitatives sur la caractérisation de la formation (réalisés à partir d'entretien semi-directifs auprès de 21 SP) et sur l'identification de l'identité professionnelle des sapeur·euse·s-pompier·ère·s (par le biais des résultats des 475 réponses au questionnaire co-construit avec le groupe de travail). Nous avons ainsi pu mettre au jour les valeurs des SP, tant au niveau de la formation qu'au niveau du métier

pour lequel iels se forment. Il s'agissait de l'étape de co-construction du référent. Dans un deuxième temps, durant la phase de la « conquête de la reconnaissance » (Marcel, 2020), nous avons croisé ces résultats avec les référentiels de formation du SDIS 31. Nous avons donc cherché à identifier dans les référentiels les compétences au service des valeurs portées par la formation et le métier de SP. Nous avons ainsi mobilisé le modèle de la gestion en définissant des compétences attendues au sein du dispositif que nous cherchions à améliorer. Enfin, pour la dernière étape, la phase d'essaimage de la RI (Marcel, 2020), nous avons opérationnalisé l'évaluation et analysé les résultats. Cette étape a mobilisé principalement le modèle de la mesure car c'est avec notre référentiel que nous avons cherché à identifier les différents critères et indicateurs définis en amont. Si les dispositifs étaient pluriels, l'objectif était bien d'observer l'adéquation entre l'attendu et l'observable.

L'évaluation ayant été entièrement mêlée à la RI, nous avons ainsi pu rendre l'évaluation collaborative, par les instances créées, et démocratique, par la mobilisation d'éléments externes aux instances lors des différentes enquêtes menées. La participation des acteur·rice·s a été déterminante dans les processus de traduction qui se sont opérés dans la mobilisation des modèles d'évaluation. Cette évaluation a ainsi permis d'apporter des changements au sein de leur ingénierie de formation et de modéliser le processus de référentialisation au travers des différentes étapes de l'évaluation en plaçant au cœur le rôle des valeurs dans l'évaluation.

V.c. Témoignage de Manon Delbreil

Notre RIE s'inscrit dans le cadre d'une thèse CIFRE menée au sein du Service Départemental d'Incendie et de Secours de la Haute-Garonne (SDIS 31) entre août 2021 et août 2024. Issue d'un partenariat entre l'UMR EFTS et le SDIS 31, amorcé par les travaux de Christelle Chauffriasse, cette démarche répond à une demande du groupement développement des compétences visant à renforcer la formation en situation de travail chez les sapeur·euse·s-pompier·ère·s. Le SDIS 31 souhaitait en effet expérimenter l'AFEST (Action de Formation En Situation de Travail) pour mieux articuler apprentissage et action dans un environnement marqué par l'urgence, une hiérarchie prégnante et la coexistence de sapeur·euse·s-pompier·ère·s professionnel·le·s et volontaires. La commande portait sur la co-construction de trois Dispositifs Expérimentaux de Formation En Situation de Travail (les DEFEST) en lien avec des réalités opérationnelles

spécifiques : le DEFEST « Immersions et Parcours » pour les chefs d'agrès en formation initiale, le DEFEST « FMPA Conduite » centré sur la spécialité conduite des engins-pompes, et le DEFEST « Sous-Officier de Garde » destiné aux managers de proximité en caserne. Ces dispositifs, conçus collectivement avec les acteurs du terrain (formateurs, accompagnateurs de proximité, agents de garde, responsables de centre), ont été co-conçus, testés, ajustés et co-évalués au fil de trois années de collaboration. Combinant observation participante et co-construction, la RIE a associé une immersion socio-ethnographique (impliquant le port de l'uniforme et la participation aux activités quotidiennes des casernes et aux interventions) à une ingénierie participative déployée dans trois groupes de travail, chacun composé d'environ dix sapeurs-pompiers représentant la diversité des grades et des fonctions. Huit entretiens semi-directifs, plusieurs *focus groups*, des captations vidéo d'auto-confrontation et des séances de co-analyse en caserne ont également alimenté le corpus empirique. Ces matériaux ont permis de rendre visibles les tensions vécues par les acteurs, entre leprescrit et le réel du travail, entre urgence opérationnelle et temps de formation, entre hiérarchie et participation. La co-évaluation, intégrée dès la conception, a constitué un véritable levier d'ingénierie. Inspirée d'une approche processuelle (Ardouin, 2023), elle a favorisé la création d'un langage commun entre formateurs, accompagnateurs et cadres sapeurs-pompiers. Dans le DEFEST SOG, par exemple, la co-construction d'un référentiel de situations professionnelles significatives (Ferron et al., 2006 ; Javier, 2022) a conduit à revisiter les modalités d'évaluation, à partir de vidéos embarquées (GoPro portées par les SOG sur leur temps de garde) et d'analyses d'auto-confrontation collectives. Cette RIE a vu émerger progressivement une pédagogie de l'opportunité**,** pensée pour un contexte de forte imprévisibilité où l'apprentissage s'entrelace à l'action. La pédagogie de l'opportunité est une approche dynamique centrée sur l'émergence et l'exploitation des occasions d'apprentissage dans des situations imprévues. Elle met l'accent sur la capacité du formateur à percevoir et à exploiter des circonstances favorables, en alliant prévisibilité et flexibilité, pour transformer l'imprévu en potentiel formatif (Ansermet, 2016). La convergence entre l'urgence de la situation (*Chronos*) et la pertinence du moment (*Kairos*) détermine si une occasion mérite d'être saisie, rejoignant les perspectives de Couloubaritsis (2016) et Fleury (2016) sur les moments propices à l'apprentissage. Elle repose sur un cycle itératif de perception, d'analyse et d'exploitation des situations vécues, articulant les temporalités du *Chronos* (temps linéaire), du *Kairos* (moment opportun) et de l'*Aiôn* (temps cyclique). Les dispositifs co-construits visent ainsi à transformer les moments du travail, souvent imprévus, en occasions d'apprentissage réflexif. Cette RI, en investissant

un tiers-espace socio-scientifique (Marcel, 2021), a permis de faire dialoguer les logiques d'action et de connaissance. La posture de la chercheuse-intervenante, impliquée et réflexive, a rendu possible la co-analyse des effets produits, tout en révélant les tensions inhérentes à toute tentative de démocratisation de la formation dans une organisation hiérarchisée. L'expérience a également mis en lumière la dimension politique de l'évaluation : comment évaluer sans figer ? Comment reconnaître la valeur formatrice de l'imprévu sans compromettre la sécurité ? Ces questionnements rejoignent une conception de l'évaluation comme espace de délibération et de reconnaissance mutuelle, où les savoirs d'expérience, souvent périphériques, trouvent une légitimité. Enfin, cette RIE a contribué à institutionnaliser, au sein du SDIS 31, des pratiques de formation situées, réflexives et participatives. Elle dévoile que, dans les milieux d'urgence, la professionnalisation ne peut être pensée sans une épistémologie de la situation, fondée sur la réflexivité, le dialogue et la co-construction des savoirs.

V.d. Témoignage de Pauline Hédacq

C'est dans le champ de la santé que nous avons mis en œuvre une démarche de RIE (Hédacq, 2023) d'une durée de quatre ans (2019/2023). Un des co-présidents de la Commission du Médicament et des Dispositifs Médicaux Stériles (CoMéDiMS) du Centre Hospitalier Universitaire (CHU) de Toulouse, nous a demandé d'accompagner une démarche d'ingénierie de formation dans le contexte de la sécurisation du circuit du médicament. Le premier enjeu de la RIE concernait le volet intervention. Elle devait contribuer à l'accompagnement de deux groupes de travail pluridisciplinaires (médecins, pharmaciens, infirmiers, sages-femmes) en charge de l'élaboration de deux dispositifs de formation. Nous avons proposé que l'ingénierie de formation ait lieu au sein du tiers-espace socio-scientifique (TESS) de la RIE (Marcel, 2021). Nous souhaitions favoriser la rencontre et la mise en dialogue des savoirs des chercheur·se·s, de la CoMéDiMS et des professionnel·le·s de santé dans un objectif de co-construction des dispositifs de formation. Le deuxième enjeu était heuristique. Nous nous sommes intéressé·e·s au processus de co-construction des savoirs au sein du TESS. Dans cette situation, nous avons posé notre regard sur la façon dont des individus pris dans des situations sociales et professionnelles différentes peuvent, dans certaines conditions, co-élaborer des savoirs contextualisés. Le troisième enjeu était épistémologique. Nous avons interrogé le rapprochement possible entre

« science » et « société ». La RIE permet par son principe participatif de prendre en compte et de valoriser les savoirs (qui peuvent présenter différentes caractéristiques) portés par les différents acteur·rice·s, quels que soient leurs statuts dans l'organisation ou dans la société. Le processus de co-élaboration des savoirs qu'elle soutient impose de repenser les rapports entre pouvoir et savoir dans une visée plus démocratique. Les acteur·rice·s engagé·e·s dans la RIE sont amené·e·s ainsi à s'émanciper, en se déplaçant d'une place à laquelle ils se situent vers une place correspondant davantage à leurs aspirations (Marcel, 2017). Dans cette démarche, l'évaluation a pris une place centrale. Pour co-construire les dispositifs de formation, nous avons opérationnalisé la démarche de référentialisation développée par Figari et al., (2014). L'objectif était de co-élaborer deux référentiels définissant respectivement un dispositif de formation pour les nouveaux et nouvelles arrivant·e·s (infirmier·ère·s et sages-femmes) sur l'administration médicamenteuse et un guide d'usage pour le rangement des armoires à pharmacie. La référentialisation, pour nous, a été mobilisée comme un opérateur de co-élaboration de savoirs. Son opérationnalisation a suivi les quatre phases d'ingénierie de formation d'Ardouin (Ardouin, 2017) : analyser, concevoir, réaliser et évaluer. Pour chacune des phases, les acteur·rice·s des groupes de travail ont défini des référents, critères et indicateurs nécessaires à l'élaboration des référentiels déterminant les dispositifs de formation à mettre en place et permettant à la CoMéDiMS de procéder à leur régulation. Du point de vue de la recherche, nous avons décrit le processus de co-élaboration des savoirs au sein du TESS et nous avons objectivé que les différents acteur·rice·s se sont déplacé·e·s d'un point de vue individuel (développer des connaissances, changer sa façon d'agir…), social (conforter ses valeurs), professionnel (augmenter son réseau, changer de poste…) et symbolique (prise de parole, valorisation, reconnaissance). La démarche de référentialisation a permis de faire émerger les savoirs des différent·e·s acteur·rice·s au sein du TESS. Elle a favorisé l'élaboration d'un diagnostic de la situation. Son principe participatif a permis aux acteur·rice·s d'être partie prenante des décisions potentialisant ainsi leurs engagements dans le projet. Les différents acteur·rice·s engagé·e·s dans la RIE ont pu penser et agir autrement dans leur monde professionnel et améliorer par la mise en place des deux dispositifs leur contexte professionnel. Nous retiendrons au travers du témoignage d'un soignant que la pluridisciplinarité a été vue comme une force : « *l'intelligence collective, c'est ça, c'est arriver à développer*

à plusieurs avec des compétences différentes et variées pour aller plus loin. » (Soignant).

VI. Conclusion

Comme nous l'évoquions en début de texte, cette contribution avait pour ambition de présenter la démarche de recherche-intervention évaluative. Nous souhaitions mettre à disposition de chercheur·se·s ou de doctorant·e·s des éléments de genèse et de théorisation de la démarche. Mais nous voulions également donner à voir sa mise en œuvre, dans sa diversité et son exigence. Les témoignages des quatre docteures ici rassemblés sont de ce point de vue éclairants, nous semble-t-il. Comme nous l'avons souligné, faire le choix d'intégrer l'évaluation dans le cadre d'une recherche de type participative, n'est pas sans risque. Cela suffirait à disqualifier le processus de recherche lui-même, dans certaines conceptions de la science. Mais à rebours, dès lors que l'on s'inscrit dans une perspective d'accompagnement du changement, l'évaluation produit un effet stimulant, elle nourrit l'adhésion de chacun·e, du fait même des enjeux décisifs qui la traverse. Elle est à ce titre un facteur d'intéressement déterminant. Elle nourrit l'intérêt des partenaires pour la recherche-intervention, à la fois concernant la dynamique intrinsèque (favorisant la participation), et du point de vue des résultats qu'elle produit (favorisant la réception). À travers les paragraphes que nous avons consacrés à l'appréhension du pouvoir au sein des RIE, on a pu percevoir qu'il ne s'agit pas ici de verser dans une vision idéaliste des choses. Mais plutôt de considérer qu'en mettant au travail ces enjeux dans un tiers espace extrêmement structuré, on peut favoriser le développement d'une réflexion collective et pacifiée. La RIE peut dès lors s'envisager comme une pratique susceptible d'accompagner le développement d'une culture démocratique.

VII. Bibliographie

Abbet, J.-P. et Suchaut, B. (2014). *Recherche en éducation et décision politique : Conditions, apports et limites de l'évaluation*. Actes du 26e colloque de l'ADMEE-Europe : Cultures et politiques de l'évaluation en éducation et en formation, Marrakech, Maroc.

https://maarifcentre.org/admee2014/images/actes/abbet-suchaut1-actes2014.pdf

Abernot, Y. (2013). Evolution d'Homo evaluatus. Dans J.-F. Marcel et H. Savy (dir.), *Evaluons, évoluons* (p. 185-196). Educagri.

Ardouin, T. (2013). *Ingénierie de formation. Analyser, Concevoir, Réaliser, Evaluer*. Dunod.

Ardouin, T. (2017). *Ingénierie de formation*. Dunod.

Ardouin, T. (2023). *Ingénierie de formation : Intégrez les nouveaux modes de formation dans votre pédagogie*. Dunod.

Arendt, H. (1967). *Essai sur la révolution*. Gallimard.

Ariza, E. (2020). *Une entreprise en changement : Évaluation d'une recherche-intervention déployée dans un service d'aide et d'accompagnement à domicile, spécialisé dans l'autisme* [Thèse de Doctorat]. Toulouse Jean Jaurès.

Audran, J. (2007). Dispositifs et situations, quelles articulations ? *Question Vives*, *8*, 11-15.

Aussel, L. (2015). Les enjeux d'une évaluation commanditée. Immersion au sein d'une recherche-intervention en sciences de l'éducation. *Formation Emploi*, *129*, 85-105.

Aussel, L. (2016). Une commande d'évaluation au sein d'une recherche-intervention. Dans J.-F. Marcel (dir.), *La recherche-intervention par les sciences de l'éducation : Accompagner le changement* (p. 107-122). Educagri.

Aussel, L. (2019). Analyse de dispositifs croisés : Vers un macro-dispositif de formation ? Le cas d'une recherche-intervention évaluative. Dans C. Ait-Ali et I. Fabre (dir.), *Le dispositif en question : Le prisme des sciences de l'éducation et de la formation et des sciences de l'information et de la communication. Détours par le macro-dispositif*. Cépaduès éditions.

Aussel, L. et Bedin, V. (2019). Analyse critique des difficultés à conduire des recherches collaboratives en évaluation. *La Revue LEeE*, *1*. https://doi.org/10.48325/rleee.001.05

Baroth, M.-F. (2012, janvier). *Didactique professionnelle et évaluation : La voix de l'émancipation*. 24e Colloque International de l'ADMEE-Europe L'évaluation des compétences en milieu scolaire et en

milieu professionnel, Luxembourg. https://admee.uni.lu/wp-content/uploads/sites/51/2018/06/ADMEE_2012_Tables-rondes_1.pdf

Bedin, V. (2009). *L'évaluation à l'université. Evaluer ou conseiller ?* Presses Universitaires de Rennes.

Bedin, V. et Broussal, D. (2015). La restitution, un enjeu de validation sociale pour la Recherche-Intervention. Dans D. Broussal, P. Ponté et V. Bedin (dir.), *Recherche-Intervention et accompagnement du changement en éducation* (p. 175-194). L'Harmattan.

Belin, E. (dir.) (2001). *Une sociologie des espaces potentiels. Logique dispositive et expérience ordinaire*. De Boeck Supérieur.

Berthelot, J.-M. (1996). *Les vertus de l'incertitude, le travail de l'analyse dans les sciences sociales*. Presses Universitaires de France.

Bertolino, I. (2024). *Contribution à la définition de la mixité sociale à l'école par l'évaluation du dispositif haut-garonnais*. Toulouse Jean Jaurès.

Bonniol, J.-J. et Vial, M. (1997). *Les modèles de l'évaluation. Textes fondateurs avec commentaires*. De Boeck.

Broussal, D., Ponté, P. et Bedin, V. (dir.) (2015). *Recherche-Intervention et accompagnement du changement en éducation*. L'Harmattan.

Chauffriasse, C. et Aussel, L. (2022). Analyse d'une référentialisation dynamique au cœur d'une évaluation collaborative d'un système de formation : Le cas d'une recherche-intervention au service départemental d'incendie et de secours de la Haute-Garonne. *La Revue LEeE*, *4*. https://revue.leee.online/index.php/info/article/view/102/410

Crozier, M. et Friedberg, E. (1977). *L'acteur et le système*. Seuil.

Dewey, J. (1967). *Logique. La théorie de l'enquête*. Presses Universitaires de France.

Dochy, F. J. R. C., Segers, M. et Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, *24*(3), 331-350.

Dupont, Y. (2014). *L'université en miettes. Servitude volontaire, lutte des places et sorcellerie*. Éditions L'échappée.

Enegrén, A. (1984). Chapitre premier—Pouvoir et violence. *Recherches politiques*, 99-114.

Figari, G. (1994). *Évaluer : Quel référentiel ?* De Boeck.

Figari, G. (2006, septembre). *Conférence de clôture : Pour une épistémologie de l'évaluation*. 19e Colloque International de l'ADMEE-Europe L'évaluation au 21e siècle : Vers de nouvelles formes, modélisations et pratiques de l'évaluation ? Luxembourg. https://admee.uni.lu/wp-content/uploads/sites/51/2018/06/ADMEE_2006_Conferences.pdf

Figari, G. et Mottier Lopez, L. (2006). *Recherche sur l'évaluation en éducation : Problématiques, méthodologies et épistémologie*. L'Harmattan.

Figari, G. et Remaud, D. (2014). *Méthodologie d'évaluation en éducation et formation ou l'enquête évaluative*. De Boeck.

Foucault, M. (1977). *Dits et écrits, volume III*. Gallimard.

Foucault, M. (2009). *Le corps utopique. Les hétérotopies*. Lignes.

Freire, P. (1974). *Pédagogie des opprimés*. François Maspero.

Hadji, C. (2014). L'évaluation est-elle condamnée à être une calamité sociale ? Dans P. Maubant, D. Groux et L. Roger (dir.), *Cultures de l'évaluation et dérives évaluatives* (p. 181-210). L'Harmattan.

Hédacq, P. (2023). *Co-élaborer des savoirs pour une ingénierie de formation : De la référentialisation à l'émancipation. Le cas d'une Recherche-Intervention conduite dans le champ de la santé*. Toulouse Jean Jaurès.

Jorro, A. (2007). *Evaluation et développement professionnel*. L'Harmattan.

Joshua, S. (1996). Le concept de transposition didactique n'est-il propre qu'aux mathématiques ? Dans C. Raisky et M. Caillot (dir.), *Au-delà des didactiques, le didactique. Débats autour de concepts fédérateurs* (p. 61-73). De Boeck.

La Fontaine, J. (1874). *Fables*. Pinot.

Lartiguet, P., Broussal, D. et Saint-Jean, M. (2022). Recherche-Intervention et formation : Dynamiques participatives pour accompagner le changement « partenariat en santé ». *Recherche & Formation*, *99*, 121-140.

Lecointe, M. (1997). *Les enjeux de l'évaluation*. L'Harmattan.

Marcel, J.-F. (2012). Contribution à une ingénierie de la commande. *Transformations*, *8*, 101-120.

Marcel, J.-F. (2014). *Lycées agricoles en changement*. Educagri.

Marcel, J.-F. (dir.) (2015). *La recherche-intervention par les Sciences de l'éducation—Accompagner le changement*. Educagri.

Marcel, J.-F. (2017). La recherche-intervention comme acte politique et émancipateur. Dans J.-F. Marcel et D. Broussal (dir.), *Émancipation et recherche en éducation* (p. 397-422). Editions du Croquant.

Marcel, J.-F. (2020). Fonctions de la recherche et participation : Une épistémo-compatibilité dans le cas de la recherche-intervention. *Questions Vives*, *33*. https://journals.openedition.org/questionsvives/4691

Marcel, J.-F. (2021). Un dispositif, un tiers-espace et des médiations. Le tiers-espace socio-scientifique dans la recherche-intervention. *Sciences de la société*, *107*. http://journals.openedition.org/sds/12659

Marcel, J.-F. et Bedin, V. (2018). Contribution à l'élaboration d'un dispositif d'évaluation de la Recherche-Intervention. *Phronesis*, *7*(1), 79-91.

Marcel, J.-F. et Broussal, D. (2017). *Emancipation et recherche en éducation*. Editions du Croquant.

Marcel, J.-F. et Broussal, D. (2022). La recherche-intervention : Une démarche pour accompagner le changement. Dans B. Albero et J. Thievenaz (dir.), *Traité de méthodologie de la recherche en sciences de l'éducation et de la formation, Tome 1* (p. 247-258). Raison et Passions.

Marcel, J.-F. et Gremion, C. (2025). *Évancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation*. Cépaduès éditions.

Miller, J.-A. et Milner, J.-C. (2004). *Voulez-vous être évalué ? Entretiens sur une machine d'imposture*. Grasset.

Mottier Lopez, L. et Figari, G. (2012). *Modélisations de l'évaluation en éducation : Questionnements épistémologiques*. De Boeck.

Nouchi, F. (1993, décembre 7). Entretien avec Pierre Bourdieu. *Le Monde*.

Peeters, H. et Charlier, P. (1999). Contributions à une théorie du dispositif. *Hermès*, *25*, 15-23.

Quelquejeu, B. (2001). La nature du pouvoir selon Hannah Arendt. *Revue des sciences philosophiques et théologiques*, *85*(3), 511-527.

Stengers, I. (2002). *Sciences et pouvoir. La démocratie face à la technoscience*. La Découverte.

Stengers, I. (2019). *Résister au désastre : Dialogue avec Marin Schaffner*. Wildproject Editions.

Stengers, I. et Schlanger, J. (dir.) (1989). *Les concepts scientifiques. Invention et pouvoir*. La Découverte.

Younès, N. et Faidit, C. (2020). Dynamiques évaluatives en jeu dans une recherche collaborative avec des enseignants et des enseignantes du second degré en France. *Mesure et Evaluation en Education*, *43*(1), 67-94.

Chapitre 9. Biais du jugement évaluatif en contexte scolaire

Condensé de 100 ans de recherche en docimologie

Dylan Dachet, Ariane Baye, Dominique Lafontaine et Marc Demeuse

I. Introduction

Notre société fait de la lutte contre les inégalités son fer de lance. Or, dans la lignée des travaux sociologiques fondateurs de Bourdieu (Bourdieu et Passeron, 1964 ; Bourdieu et Passeron, 1970), les enquêtes internationales chapeautées par l'OCDE (Monseur et Baye, 2015) et l'IEA (Lafontaine et al., 2019) ont montré que les élèves n'étaient pas égaux face à l'École. Si ces résultats ont exhorté les systèmes éducatifs à réduire les inégalités, la question de leur origine continue à susciter des débats. Parmi les sources de ces inégalités, l'évaluation scolaire occupe une place particulière. D'une part, les biais engendrés par certaines pratiques évaluatives sont à l'origine d'inégalités de traitement accentuant les inégalités culturelles, sociales et économiques. D'autre part, la tradition séculaire de recherche sur l'évaluation a été fertile et a mis en évidence que, sous certaines formes, l'évaluation pouvait éviter ces écueils. Dès lors, selon les usages pédagogiques et sociaux qui en sont faits, l'évaluation peut tant renforcer que réduire les inégalités scolaires. Ces constats relèvent d'une discipline scientifique connue sous le nom de docimologie - nom que l'on doit à Henri Piéron et qui résulte de l'agrégation de deux mots grecs : « dokimé, épreuve, et logos, science » (Martin, 2002, p. 177) - et ont été posés dans plusieurs disciplines. Ainsi, des inégalités de jugement ont été identifiées

chez les banquiers (Kahneman et al., 2021), les recruteurs (Bertrand et Mullainathan, 2004), les juges (Austin et Williams, 1977 ; Clancy et al., 1981) et les enseignants. Or, dans ces différents domaines, l'évaluation produit « une "vérité" sociale qui peut être lourde de conséquences pour ceux qui en font l'objet » (Bressoux et Pansu, 2003, p. 7) ; d'où la nécessité d'en faire une étude approfondie. Ce chapitre retrace l'histoire de la docimologie en partant des premières études pointant le manque d'objectivité de la notation scolaire et en concluant par les travaux récents concernant les méthodes limitant les biais de l'évaluation.

II. Émergence de la docimologie descriptive

Les études princeps à l'origine du développement de cette discipline ont été conduites au début du XX^{e} siècle. La docimologie qui s'attachait alors à décrire les « systèmes de notation, et [le] comportement des examinateurs et des examinés » (De Landsheere, 1980, p. 17) a principalement établi ce qui peut désormais sembler une évidence. La notation scolaire est sensible au bruit et aux biais. Ces phénomènes ont été constatés à partir de deux expériences fameuses. D'une part, lors de l'expérience conduite par Laugier et Dagmar (1936), six correcteurs ont attribué des notes sur une échelle de 20 points à 100 copies. Des écarts importants ont été observés dans les notes attribuées, allant de 8 points en physique à 13 points en composition française. Cette dispersion des notes attribuées à une même production par différents évaluateurs est une illustration du bruit. D'autre part, lors de l'expérience conduite par Gjorgjevski (cité par Rot et Bujas, 1959, p. 20), 15 copies qui avaient obtenu unanimement l'appréciation « bien » parmi un corpus de 100 copies évaluées de façon indépendante par cinq enseignant.e.s, ont été soumises à l'évaluation de quatre autres enseignant.e.s. Là où les cinq enseignant.e.s avaient jugé ces copies comme homogènes, les quatre enseignant.e.s qui interviennent ensuite ont décelé une variabilité importante. Cet écart systématique entre les appréciations formulées par les différents évaluateurs dans ces deux conditions illustre le phénomène de biais. Dans ce cas précis, une hypothèse explicative de la variabilité peut être formulée : « le changement de composition du groupe [de copies] a provoqué un changement de standard des correcteurs, qui [leur ont attribué] différentes notes allant d'excellent à médiocre ».

III. Développement d'une docimologie critique

Après ces premiers constats, la docimologie a basculé dans un versant plus critique au milieu du XX^{e} siècle. Elle fut alors qualifiée comme telle dès lors qu'elle se centrait sur « la mise en évidence des problèmes et injustices » (Leclercq et al., 2004, p. 275) produits par les systèmes d'évaluation. Pour y parvenir, elle a mobilisé des protocoles expérimentaux. Ces plans de recherche reposent sur la comparaison entre un groupe expérimental et un groupe témoin différant l'un de l'autre en fonction de la manipulation de la variable indépendante dont on cherche à étudier l'effet causal sur la variable dépendante (Dachet et Baye, 2021). Parmi les études expérimentales, on peut distinguer les essais contrôlés randomisés, où l'assignation des participant.e.s aux groupes suit une procédure strictement aléatoire, des designs quasi expérimentaux où les groupes sont constitués sans manipulation du chercheur (Slavin, 2007). Lors de ces études, l'objectif de la recherche est délibérément caché aux enseignant.e.s afin d'éviter une modification de leurs attitudes. En outre, les productions utilisées impliquent généralement une évaluation subjective, c'est-à-dire à partir de critères de qualité dont l'appréciation peut varier d'un évaluateur à l'autre (rédaction, réponses longues…).

Dans ce contexte, Bonniol (1972) a réalisé une étude expérimentale où deux groupes de 9 correcteurs ont évalué des devoirs. Ces groupes se distinguent sur un seul point : les correcteurs du premier ont reçu les devoirs à corriger dans un ordre donné alors que les correcteurs du second les ont reçus dans l'ordre inverse. Les appréciations fournies dans les deux groupes différaient de telle façon à ce que les divergences observées soient « imputables aux deux ordres de corrections […] » (Bonniol, 1973, p. 203). L'auteur a ensuite introduit des ancres au sein des séries de copies, c'est-à-dire des stimuli jouant le rôle de référence (Tversky et Kahneman, 1974) pour les correcteurs. Il a, ainsi, placé dans chaque série une copie dont les propriétés sont connues et dont l'appréciation souvent extrême (excellente/médiocre) peut servir de référence aux correcteurs. Lors de deux expériences, Bonniol (1972) a assigné aléatoirement des professeur.e.s à sept groupes dans lesquels ils ont évalué des séries de copies qui ne différaient que par le nombre, le niveau et la place des ancres introduites. Les résultats ont permis de confirmer l'hypothèse selon laquelle l'évaluation d'une copie est influencée par le niveau des copies précédentes. Ce faisant, il a mis en évidence l'effet d'ordre ou de séquence (Leclercq, 1999). Cet effet se manifeste lorsque l'évaluation des performances d'un élève est influencée par la « qualité » des copies

antérieurement évaluées (*Between-Subject*). Ce biais est contrastif (Humphrey-Murto et al., 2021), puisque l'évaluateur modifie inconsciemment son appréciation en sorte qu'elle s'éloigne des appréciations antérieures.

En parallèle, la docimologie a également mis en exergue un effet de contagion ; aussi connu sous le nom d'effet de réputation (Findlay et Ste-Marie, 2004), de stéréotypie (Leclercq, 1999) ou d'inertie (Leclercq et al., 2004). Cet effet se manifeste lorsqu'un.e enseignant.e évalue successivement plusieurs tâches réalisées par un même élève (*Within-Subject*), par exemple les différentes questions d'un examen, et que sa réussite aux premières tâches contamine l'appréciation des tâches suivantes. La recherche expérimentale (Shaw et al., 2021) a mis en évidence que cet effet est assimilatif (Humphrey-Murto et al., 2021), puisque l'évaluateur biaise inconsciemment son appréciation pour qu'elle se rapproche de ses appréciations antérieures. Par exemple, Malouff et al. (2013) ont assigné aléatoirement 126 enseignant.e.s en psychologie de 23 universités australiennes à deux conditions expérimentales. Dans la première, 65 enseignant.e.s ont évalué l'analyse écrite d'un cas clinique réalisée par une étudiante après avoir évalué une vidéo durant laquelle cette étudiante répond médiocrement à un examen oral tout en présentant une image négative d'elle-même (tenue et apparence négligées). Dans la seconde, 61 enseignant.e.s ont évalué cette même analyse écrite, mais après avoir évalué une vidéo durant laquelle l'étudiante répond de façon habile à l'examen oral tout en présentant une image positive d'elle-même (tenue et apparence soignées). Les enseignant.e.s ayant observé l'étudiante dans des circonstances favorables ont évalué plus positivement sa réponse écrite que les enseignant.e.s qui l'ont préalablement observée dans des circonstances moins favorables.

Enfin, les travaux menés par Asch (1946) ont fourni les premières preuves empiriques de l'effet de halo. Ce chercheur a étudié la formation des impressions en comparant des descriptions écrites par des étudiant.e.s concernant des personnes pour lesquelles on leur fournissait une liste d'adjectifs. Il a mis en évidence que les descriptions étaient plus positives lorsque la liste commençait par des éléments positifs. Ainsi, nous organisons nos impressions afin de fournir une description cohérente, ou homogène (Sanrey et al., 2021), de la personne évaluée. Notre jugement d'un individu est ainsi influencé par la première impression que nous nous faisons à partir de l'une de ses caractéristiques. L'évaluation scolaire peut ainsi être influencée par l'attitude d'un élève, les caractéristiques formelles de son travail… Étant donné la quantité d'études menées récemment à ce sujet, une section entière est consacrée à la description de cet effet.

L'effet d'ordre, l'effet de contagion et l'effet de halo peuvent être regroupés sous le concept générique d'effet de contexte (Palmer et Gore, 2014).

IV. Docimologie du jugement au sein des classes

Au début des années 1990, les docimologues ont profité de la généralisation des évaluations externes[1] ou standardisées au sein des systèmes éducatifs pour comparer les résultats obtenus par les élèves à ces épreuves externes aux résultats qu'ils ont obtenus aux épreuves internes (c'est-à-dire aux épreuves construites et conduites par leur enseignant). En recourant à des modèles multi-niveaux, la docimologie s'est alors intéressée aux inégalités produites par le jugement évaluatif des enseignant.e.s en contexte de classe réel. Grâce à ces modèles, les chercheur.e.s ont analysé les données récoltées en prenant en considération leur structure hiérarchique : les élèves sont nichés dans des classes, elles-mêmes nichées dans des écoles. Ces modèles ont permis d'isoler l'influence des facteurs individuels et des facteurs liés à l'école dans l'évaluation scolaire.

En Belgique, la recherche APER[2] (Grisay, 1984) est la première à avoir comparé les résultats obtenus par 1 503 élèves issus de 53 écoles de la Province de Liège à un examen interne, conçu et corrigé par leur enseignant.e aux résultats obtenus par ces mêmes élèves à une épreuve externe. Cette comparaison a montré que la variance des scores des élèves au test externe était systématiquement plus faible que la variance de leurs scores à l'évaluation interne. « Tout se passe comme si l'évaluation pratiquée par les maîtres produisait une amplification de l'hétérogénéité des performances des élèves » (Crahay, 2019, p. 366). Ces résultats ont démontré l'existence de l'effet Posthumus, en référence à l'auteur qui fut le premier à le constater[3]. Parfois appelé « constante macabre » (Antibi,

[1] Une épreuve externe est un test mesurant la performance des élèves dont la conception, la mise en œuvre et la correction sont effectuées par des personnes extérieures aux équipes éducatives (FW-B, 2006, p. 1). En Belgique francophone, seul le Certificat d'Études de Base (CEB) répond à ce critère.

[2] Amélioration des Pratiques d'Évaluation-bilan en vue d'une réduction des Redoublements.

[3] Texte initial de Kees Posthumus, paru dans *De Guids* n° 104 de 1940 sous le titre *Middelbaar onderwijs en schifting* (pp. 24-42) : https://www.dbnl.org/tekst/_gid001194001_01/_gid001194001_01_0040.php

2003), cet effet rend compte du fait que les notes des élèves d'une même classe en fin d'année scolaire épousent une distribution normale et ce, indépendamment de la distribution initiale de leurs performances. Selon de Landsheere (1980, p. 224), « l'enseignant adapte le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver, d'année en année, approximativement la même distribution (gaussienne) de notes ». Ainsi, l'enseignant.e adapte inconsciemment son niveau d'exigence et de sévérité en fonction du niveau moyen des élèves de sa classe.

La conduite d'études mesurant la sévérité des enseignant.e.s est éthiquement impossible, puisque cela impliquerait de définir un seuil de sévérité afin de comparer les enseignant.e.s entre eux. En revanche, il est possible d'étudier un « effet correcteur » à partir des écarts constatés entre les appréciations de plusieurs correcteurs sur une même production. En recourant à la théorie de la généralisabilité et au modèle de Rasch, Casanova et Demeuse (2011) ont ainsi montré des écarts de sévérité manifestes entre évaluateurs. En outre, des études ont pointé que l'hétérogénéité des critères d'évaluation est l'une des causes de la variabilité des notes observée entre les classes. Après avoir demandé à 16 professeur.e.s de français de pondérer les critères d'un examen comptant pour 60 points, Detheux-Jehin et Kellens (1992) ont constaté : (1) des divergences dans les critères utilisés, puisque par exemple le vocabulaire n'est évalué que par 11 d'entre eux ; (2) des divergences importantes dans la pondération accordée à chaque critère, puisque par exemple la pondération accordée à l'orthographe varie entre 2 et 22 points. Si ces différences peuvent relever de la liberté pédagogique des enseignant.e.s, elles ne sont pas anecdotiques pour l'élève. Par exemple, un élève qui présenterait des difficultés en grammaire risquerait de perdre 14 points avec le professeur X alors qu'il risquerait de perdre 39 points et d'être en échec avec le professeur Y.

En comparant ainsi les résultats des mêmes élèves à des épreuves externes et internes, les docimologues ont parallèlement tenté d'expliquer les divergences entre ces deux évaluations à partir des caractéristiques des élèves évalués. Il s'agit alors de designs corrélationnels de recherche (Dachet et Baye, 2021) qui ont largement mis en évidence des effets de halo lors des évaluations internes et agissant au détriment :

1. des garçons en Angleterre (Burgess et Greaves, 2009), au Chili (Contreras, 2023), au Danemark (Rangivd, 2015), aux États-Unis (Ouazad, 2008), en Israël (Lavy, 2004), en Italie (Di Liberto et al., 2022), au Portugal (Angelo, 2014 ; Angelo et Reis, 2017, 2021), en

République tchèque (Protivinsky et Munich, 2018) et en Suède (Berg et al., 2019 ; Berg et al., 2020 ; Lindahl, 2007), et ce en particulier dans les disciplines langagières ;

2. des élèves d'origine immigrée victime de stéréotype[4] négatif en Angleterre (Burgess et Greaves, 2013), au Brésil (Botelho et al., 2015), en Italie (Alesina et al., 2018), en Équateur (Drouet et al., 2023), au Danemark (Rangivd, 2015) et en Suède[5] (Hinnerich et al., 2010, 2011 et 2015) ;
3. des élèves d'origine sociale et économique défavorisée en Angleterre (Burgess et Greaves, 2009), au Danemark (Rangivd, 2015), en France (Bressoux et Pansu, 2003) ;
4. des élèves à besoins spécifiques en Angleterre (Burgess et Greaves, 2009) ;
5. des élèves ayant redoublé en France (Bressoux et Pansu, 2003) ;
6. des élèves en surpoids aux États-Unis (Rouse et Hunziker, 2020 ; Zavodny, 2013).

En parallèle, d'autres études corrélationnelles se sont attelées à étudier l'effet des caractéristiques personnelles de l'enseignant.e. Ces études ont mis en évidence que les élèves évalués par un.e enseignant.e d'une origine ethnique identique à la leur bénéficiaient d'une surestimation de leurs performances. Dee (2004) a mis en évidence cet effet à partir des données

[4] Un stéréotype est une « croyance partagée par un groupe à propos des caractéristiques de personnalité et de comportement de l'ensemble d'un autre groupe » (Leyens et Yserbyt, 2017, p. 50).

[5] Les travaux de Lindahl (2007) montrent un effet inverse : les élèves d'origine étrangère bénéficient d'une surestimation de leurs performances lors de la prise de décision relative à l'obtention du School Leaving Certificate. Nous faisons l'hypothèse que ces résultats sont dus à l'enjeu très particulier de cette évaluation qui conditionne l'avenir des élèves. Face à cet enjeu, les enseignant.e.s manifesteraient de façon générale une plus grande indulgence afin d'éviter aux élèves une situation d'échec dramatique pour leur avenir (Lafontaine et Monseur, 2009). Ce faisant, et puisque ces publics précarisés ont déjà en moyenne des résultats plus faibles, l'obtention du certificat nécessite pour eux une surestimation plus importante de leurs performances que celle qu'on pourrait observer au sein d'une population moins précarisée. Dès lors, les différences observées ne seraient pas la conséquence d'un effet de halo, mais seraient dues au fait que les surévaluations des enseignants touchent essentiellement les élèves en échec plus nombreux dans les populations immigrées.

du Projet STAR où les élèves du Tennessee ont été assignés aléatoirement à des classes et à des enseignant.e.s spécifiques. Ouazad (2008) a mis en évidence un effet similaire à partir des données de la *Early Childhood Longitudinal Study* récoltées aux États-Unis auprès d'un échantillon représentatif de 20'000 élèves de l'enseignement maternel et suivis jusqu'en 2e secondaire. En outre, Dee (2005) a montré que cet effet touchait également l'évaluation comportementale des élèves. Lors de la *National Education Longitudinal Study*, le comportement de 24'599 étudiants a été évalué par deux enseignants distincts. Les analyses réalisées montrent que les élèves ont plus de risques d'être qualifiés de perturbateurs ou d'inattentifs ou encore, d'être considérés comme peu assidus lorsqu'ils sont évalués par un.e enseignant.e d'une origine ethnique différente de la leur.

Sur la question de l'appariement de genre entre élève et enseignant, les données sont moins consistantes. Alors que Dee (2005) montre que les élèves ont tendance à être évalués plus négativement par un.e enseignant.e du genre opposé, d'autres études (Chouinard, 2002) ne montrent aucun effet du genre de l'enseignant sur l'évaluation portée.

V. Études et méta-analyse sur l'effet de halo

Les effets de halo mis en évidence au sein de ces études corrélationnelles indiquent clairement que l'évaluation scolaire produit des discriminations. Ces designs de recherche ne sont cependant pas en capacité de déterminer l'effet de halo propre et indépendant de chacune des caractéristiques personnelles étudiées. À cette fin, des études expérimentales de plus en plus rigoureuses ont été conduites au début du XXIe siècle, poussées par un mouvement prônant le développement de connaissances causales en éducation (Slavin, 2002). Ainsi, en parallèle de la massification de la publication scientifique (Bastian et al., 2010), le nombre de recherches publiées dans le domaine de la docimologie (Baye et Dachet, 2022 ; Hanna et Linden, 2009 ; Nieminen et al., 2013 ; Rauw et al., 2023…) a suivi une courbe exponentielle analogue à ce qui a été observé plus globalement dans les sciences de l'éducation (Connolly et al., 2018). La communauté scientifique a donc dû se doter de méthodes rigoureuses pour synthétiser la pléthore de recherches disponibles. Dans ce cadre, la méta-analyse a été développée par Glass (1976) afin de synthétiser en un indice statistique moyen les résultats issus de toutes les études réalisées sur une même question de recherche (Borenstein et al., 2009).

En 2016, Malouff et Thorsteinsson ont ainsi réalisé une méta-analyse des 20 études expérimentales conduites afin de déterminer les biais ayant un impact sur l'évaluation des performances scolaires des élèves. Cette méta-analyse conclut que « les étudiants qui sont membres d'un groupe contre lequel il existe un préjugé auront tendance à recevoir de moins bonnes notes que les étudiants qui ne font pas partie de ce groupe » (2016, p. 5). La littérature scientifique synthétisée semble avoir essentiellement traité la question des effets de halo impactant les élèves atteints de déficiences intellectuelles (8 études) ou issus de l'immigration (7 études). Au-delà de l'effet moyen, les auteurs pointent des variances importantes entre les études semblant indiquer que l'ampleur du phénomène peut varier d'une étude à l'autre.

Sur la question de l'effet de halo affectant les élèves d'origine immigrée, les études de Sprietsma (2013) et Van Ewijk (2011) offrent une hypothèse explicative assez séduisante. Ces chercheurs ont montré que l'effet de halo était plus fort lorsque l'inférence évaluative exigée des enseignant.e.s devait aboutir à l'attribution de places inégales dans un contexte méritocratique. Ainsi, lorsqu'on demande aux enseignant.e.s de recommander pour des étudiant.e.s une orientation vers une filière d'enseignement plus ou moins prestigieuse, les stéréotypes peuvent exercer une influence prépondérante. À titre d'exemple, alors que chaque enseignant.e a été amené.e à évaluer les mêmes copies tantôt attribuées à des élèves d'origine immigrée, tantôt à des élèves d'origine native, Sprietsma (2013) a montré que les enseignant.e.s ont plus fréquemment recommandé la filière *Gymnasium* (préparatoire à l'Université) aux élèves allemands et la filière *Hauptschule* (professionnelle) aux élèves d'origine turque.

Une hypothèse alternative pour expliquer la variabilité des effets de halo concerne l'interaction entre la qualité (c'est-à-dire le niveau de performance démontré par l'élève) de la copie et le stéréotype qui peut affecter l'élève évalué. Casanova et Demeuse (2017) ont ainsi montré que la sévérité des correcteurs pouvait s'exercer de façon non homogène en fonction du niveau de performance de la copie évaluée. De leur côté, Lafontaine et Monseur (2009) ont demandé à 48 professeur.e.s de mathématiques d'évaluer quatre copies fictives de niveaux différents : une copie très faible, une copie assez faible, une copie assez bonne et une très bonne copie. Les mêmes copies ont été attribuées à des prénoms soit clairement masculins, soit clairement féminins. Les copies faibles attribuées à des filles ont eu tendance à bénéficier d'une forme de clémence de la part des professeurs se traduisant par une surestimation du score qui leur a été attribué comparativement au score obtenu par les garçons.

Parallèlement, les professeurs ont eu tendance à mieux noter les bonnes copies attribuées à des garçons comparativement à celles attribuées à des filles. En complément de l'étude expérimentale, Lafontaine et Monseur (2009) ont mené une étude en milieu naturel. Ils ont analysé les données issues de PISA 2000 où on disposait à la fois des scores des élèves à l'épreuve externe PISA et des notes de leur enseignant.e de mathématiques. Les résultats sont tout à fait cohérents avec ceux de l'étude expérimentale. A performances égales dans PISA, les garçons peu performants dans PISA étaient sous-notés par leurs enseignant.e.s par rapport aux filles de même niveau. De leur côté, les filles très performantes dans PISA étaient sous-notées par leurs enseignants, comparativement aux garçons de même niveau.

VI. Apports de la psychologie

Durant toutes ces années, la docimologie s'est développée en parallèle des sciences psychologiques. Si les deux disciplines s'accordent sur l'existence des biais, la psychologie indique qu'ils ne sont que la face visible de l'iceberg, puisqu'ils traduisent les attentes différenciées des enseignant.e.s concernant leurs élèves. En effet, les méta-analyses conduites par Tenenbaum et Ruck (2007) indiquent que les enseignant.e.s font preuve d'attentes différenciées concernant la performance de leurs élèves notamment en fonction de l'origine de ces derniers. Les enseignant.e.s développent des attentes plus favorables concernant les élèves d'origine asiatique ou européo-américaine et des attentes plus défavorables concernant les élèves d'origine hispanique ou afro-américaine. Figlio (2005) a mis en évidence un même effet au détriment des élèves défavorisés.

S'ils sont le résultat des attentes différenciées des enseignant.e.s, les biais évaluatifs pourraient résulter d'un phénomène psychologique plus vaste. La recherche en psychologie (Darley et Gross, 1983) a mis en évidence l'existence d'une tendance de l'être humain à rechercher des informations qui confirment ses croyances. Cette tendance, qui porte le nom de biais de confirmation, conduit par exemple, les personnes favorables à la peine de mort (Lord et al., 1979) et les climatosceptiques (Sunstein et al., 2017) à accorder plus de crédit aux arguments fallacieux allant dans le même sens que leurs croyances (Romainville, 2023). Cherchant à confirmer leurs attentes, les enseignant.e.s seraient alors plus enclin.e.s à pénaliser des erreurs lors de l'évaluation d'élèves dont les performances s'écartent de ce qui est attendu en fonction des stéréotypes affectant leur groupe (par

exemple, un garçon faible en mathématiques). Ces phénomènes peuvent produire des effets délétères pour certains élèves, notamment pour ceux "qui font mieux qu'attendu". En effet, la recherche a montré que, soumis à des attentes spécifiques concernant son devenir scolaire, un élève pouvait tendre à confirmer les prédictions d'échec formulées à son égard (Raynal et Rieunier, 2014). On parle alors de prophétie autoréalisatrice ou d'effet Pygmalion (Rosenthal et Jacobson, 1971). Ainsi, les élèves dévalorisés par les biais de l'évaluation et faisant l'objet d'attentes moindres de la part des enseignant.e.s pourraient tendre inconsciemment à confirmer les attentes négatives dont ils sont porteurs et pourraient réellement obtenir de moins bonnes performances. Si cette diminution des performances résulte de processus psychologiques subtils, elle peut également découler d'une diminution des efforts (Jussim et Harber, 2005) et de l'investissement dans les tâches scolaires (Fresson et Dardenne, 2021) des élèves concernés.

Par ailleurs, ces élèves pourraient également subir les effets négatifs de de la menace du stéréotype (Steele, 1997). Ce phénomène se manifeste lorsqu'un individu, conscient du stéréotype négatif porté par son groupe, confirme ce stéréotype dans une situation où il s'applique. Ainsi, les élèves afro-américains sous-performent à un test lorsque celui-ci est présenté comme évaluant leurs capacités plutôt que lorsqu'il est présenté comme un simple exercice (Steele et Aronson, 1995). En outre, des personnes souffrant d'atteintes neuronales démontrent des performances inférieures à une épreuve de concentration et de mémoire lorsqu'elle est présentée comme un test diagnostique plutôt que comme un test de vision et d'audition (Fresson et al., 2018).

Dans un contexte mêlant attentes différenciées, discriminations systémiques, effet Pygmalion et menace du stéréotype, les conséquences à long terme pour les élèves dévalorisés s'avèrent inquiétantes. Traitant essentiellement de la problématique des différences de genre, Mechtenberg (2009) a montré que les filles victimes de biais en mathématiques et en sciences sont moins nombreuses parmi les percentiles des élèves les plus performants dans ces deux domaines. Ce constat a été confirmé par Baye et Monseur (2016) dans leur réanalyse des données des enquêtes internationales PISA, PIRLS et TIMSS réalisées entre 1995 et 2012. Moins nombreuses dans les hauts niveaux de la distribution des performances en mathématiques et en sciences, les filles s'inscrivent dès lors moins fréquemment dans des filières universitaires liées au domaine du STEM[6] et obtiennent *in fine* des salaires moindres sur le marché du

[6] *Science, Technology, Engineering and Mathematics.*

travail (Jaegers et al., 2023). Martinot et al. (2025) ont d'ailleurs montré que des inégalités de genre pouvaient se manifester dès les premiers mois de scolarité. Bien que des hypothèses rationalistes (choix stratégique par anticipation des obstacles, par exemple) puissent également expliquer ces différences (Dubet, 2010), elles constituent également le signe tangible que les biais évaluatifs peuvent être des vecteurs du déterminisme social et d'inégalités liées au genre.

VII. Peut-on lutter contre les biais évaluatifs ?

Il reste à aborder une question centrale : sommes-nous en mesure de nous départir de ces biais ? Lors de quatre études expérimentales, Autin et al. (2019) et Batruch et al. (2019) répondent par l'affirmative à cette question. Ces derniers ont mis en évidence que le contexte sélectif dans lequel l'évaluation scolaire est conduite est pour beaucoup dans l'ampleur observée des effets de halo. Pour ce faire, ces chercheurs ont eu recours à une méthodologie classique : des correcteurs ont corrigé des dictées identiques attribuées à des élèves économiquement favorisés ou défavorisés. Dans chaque condition, ils ont introduit une seconde variable indépendante : une présentation du contexte évaluatif. À certains correcteurs, les chercheurs ont indiqué que l'évaluation avait pour but de sélectionner les meilleurs élèves. Dans ce contexte sélectif, ils ont demandé aux correcteurs de souligner les erreurs repérées et d'attribuer un score. Aux autres correcteurs, les chercheurs ont indiqué que l'évaluation avait pour but d'aider les élèves à s'améliorer en identifiant des remédiations adaptées. Dans ce contexte formatif, ils ont demandé aux correcteurs de souligner les erreurs repérées et d'écrire un commentaire. Alors qu'un effet de halo est présent en contexte sélectif, il disparait en contexte formatif. Ainsi, en modifiant le contexte et la visée de l'évaluation, l'enseignant est en mesure de diminuer l'impact des biais. Ces résultats sont d'ailleurs cohérents avec une méta-analyse récente démontrant les effets positifs de l'évaluation formative (Lee et al., 2021) qui peut s'ériger en réel soutien à l'apprentissage (Lafontaine et Toczek-Capelle, 2023).

Si cette solution prend tout son sens dans une vision éducative de l'École, il n'en reste pas moins inévitable qu'un système éducatif organise, ne serait-ce qu'occasionnellement et à des moments charnières, des évaluations certificatives (obtention d'un diplôme…), voire des évaluations pronostiques (examen d'entrée…). Dans ces cas, des solutions techniques telles que la correction anonymisée, la correction en double aveugle et la correction *question par question* doivent être préconisées.

Souvent coûteuses, ces solutions restent les meilleures dont nous disposions à l'heure actuelle pour tenter de faire de la correction en contexte évaluatif un processus juste et équitable. Dans certains domaines, des méthodologies innovantes sont mises en œuvre afin de garantir l'anonymat des personnes évaluées. Ainsi, dans le domaine de la musique par exemple, des auvents sont parfois utilisés lors des auditions afin de dissimuler les caractéristiques personnelles des artistes, tout en permettant au jury d'observer leur position, leurs mouvements et, partant, d'évaluer leur maîtrise technique.

En parallèle de ces solutions techniques, des auteur.e.s plaident en faveur d'une (in)formation des enseignant.e.s sur les biais de l'évaluation, pour diminuer leur degré de certitude quant à leur jugement. En effet, Sanrey et al. (2021) ont étudié l'effet modérateur du degré de certitude de l'enseignant envers le jugement qu'il pose sur l'ampleur de l'effet de halo. Ces chercheur.e.s ont montré que cet effet est d'autant plus fort que l'enseignant est sûr de son jugement. Ainsi, « les enseignant.e.s qui sont sûr.e.s de leur jugement peuvent estimer qu'ils n'ont pas besoin d'analyser attentivement chacune des réponses de leurs élèves lorsqu'ils doivent porter un jugement » (Sanrey et al., 2021, p. 671) et sont dès lors plus enclins à être victimes des effets de halo.

VIII. Conclusion

Ce chapitre dresse un état des lieux, sans prétendre à l'exhaustivité, de la recherche en docimologie. Il met en évidence un constat aussi frappant qu'inquiétant. Malgré un regain d'intérêt récent pour les questions relatives aux discriminations et les évolutions observées à cet égard au sein de la société, les conclusions formulées par les recherches contemporaines ne remettent pas en cause les conclusions établies depuis plusieurs décennies. En effet, l'accumulation des preuves issues de la recherche corrélationnelle et expérimentale est sans appel. Le jugement évaluatif des enseignant.e.s est biaisé par des phénomènes inconscients qui rendent l'évaluation subjective et souvent peu équitable. En outre, ces mécanismes inconscients à l'œuvre chez les enseignant.e.s d'hier semblent encore à l'œuvre de nos jours chez les enseignant.e.s. Si ces phénomènes sont à l'origine d'une augmentation de la variance intra- et inter-évaluateurs, ils sont surtout à l'origine de discriminations systémiques qui désavantagent certains élèves ; ces mêmes élèves qui sont souvent en situation précaire à l'École.

L'analyse des études démontrant l'existence des effets de halo est éclairante à ce sujet. À niveau de performance équivalent, les élèves objets d'attentes faibles ou victimes de stéréotypes négatifs voient leurs performances sous-estimées. Bien que les biais de l'évaluation ne se traduisent pas toujours en des différences de grande ampleur, la répétition de ces discriminations tout au long de la scolarité peut avoir des effets dramatiques, dépassant le cadre de l'évaluation. Entre effet Pygmalion, intériorisation et menace du stéréotype, tout porte à croire que ces discriminations peuvent influer sur le parcours scolaire et de vie des élèves qui en sont victimes. Ces derniers pourraient ainsi se désinvestir de certaines matières scolaires, perdre confiance dans leurs capacités, déserter certaines filières d'enseignement et ne pas bénéficier des mêmes chances d'émancipation sociale.

Au-delà d'une description des écueils de l'évaluation, la docimologie formule également des recommandations visant à améliorer la qualité et la validité de l'évaluation scolaire. Parmi ces recommandations, le rôle de la visée de l'évaluation a été plus largement examiné. En effet, les biais évaluatifs seraient le signe tangible des discriminations existantes dans les systèmes éducatifs structurés autour d'idéaux sélectifs. Lorsqu'on demande à des correcteurs d'évaluer des copies dans une perspective plus formative, ces biais s'atténuent, voire disparaissent. Cette perspective formative de l'évaluation scolaire n'est pas sans rappeler la pédagogie de la maîtrise que Bloom (1968) appelait de ses vœux. Plus de 50 ans après ses travaux, l'efficacité de l'évaluation formative comme soutien à l'apprentissage (Fagnant et al., 2014) des élèves est largement démontrée, à condition qu'elle soit correctement mise en oeuvre.

Enfin, le degré de certitude qu'un évaluateur a envers son jugement influence également celui-ci. Moins un évaluateur est certain de son évaluation, moins il semble affecté par les biais. Si nous sommes contraints, en raison de la subjectivité de notre jugement, à cultiver une forme de doute systémique sur la qualité des évaluations qui en résultent, nous savons, en revanche, que les renforcements positifs peuvent agir comme le moteur de l'apprentissage des élèves (Royer et al., 2018). Face à l'incertitude de notre jugement, il nous semblerait dès lors opportun de privilégier l'évaluation formative constante des élèves ainsi que la valorisation régulière de leurs qualités tout en mettant en œuvre, de façon occasionnelle, des évaluations sommatives ou pronostiques en recourant, *a minima* et systématiquement, à l'anonymisation des élèves.

IX. Bibliographie

Alesina, A., Carlana, M., La Ferrara, E. et Pinotti, P. (2018). Revealing stereotypes: Evidence from immigrants in schools. *NBER Working Paper, 25333*, 1-48. ttps://www.nber.org/system/files/working_papers/w25333/w25333.pdf

Angelo, C. (2014). *Is there a bias towards girls in non anonymous evaluation? Economics of Education*. https://2014.economicsofeducation.com/user/pdfsesiones/166.pdf

Angelo, C. et Reis, A. B. (2017). Gender gaps in different grading systems. *Working Paper, 625*, 1-33.

Angelo, C. et Reis, A. B. (2021). Gender gaps in different grading systems. *Education Economics, 29*(1), 105-119. https://doi.org/10.1080/09645292.2020.1853681

Antibi, A. (2003). *La Constante Macabre : comment a-t-on découragé des générations d'élèves ?* Math'Adore.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*(3), 258–290. https://doi.org/10.1037/h0055756

Austin, W.G. et Williams, T.A. (1977). A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity. *Journal of Criminal Law & Criminology, 68*, 306. https://doi.org/10.2307/1142852

Autin, F., Batruch, A. et Butera, F. (2019). The function of selection of assesment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology, 111*, 717-735. https://doi.org/10.1037/edu0000307

Bastian, H., Glasziou, P. et Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS medicine, 7*(9), e1000326. https://doi.org/10.1371/journal.pmed.1000326

Batruch, A., Autin, F. et Butera, F. (2019). The paradoxical role of meritocratic selection in the perpetuation of social inequalities at school. Dans J. Jetten et K. Peters (dir.), *The social psychology of inequality*. Springer.

Baye, A. et Dachet, D. (2022). *A randomized controlled trial on the « COVID generation »: Were students' assessments biased during the health crisis?* [Paper presentation]. BAPS 2022, Leuven, Belgium. https://hdl.handle.net/2268/291711

Baye, A. et Monseur, C. (2016). Gender differences in variability and extreme scores in an international contexte. *Large-scale Assessments in Education, 4*, 1-16. https://doi.org/10.1186/s40536-015-0015-x Berg, P., Palmgren, O. et Tyrefors, B. (2019). Gender grading bias in junior high school mathematics. *Working Paper, 1263*, 1-11. http://hdl.handle.net/10419.210904

Berg, P., Palmgren, O. et Tyrefors, B. (2020). Gender grading bias in junior high school mathematics. *Applied Econmics Letter, 27*(11), 915-919. https://doi.org/10.1080/13504851.2019.1646862

Bertrand, M. et Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*, 991-1013. https://doi.org/10.3386/w9873

Bloom, B. S. (1968). Learning for mastery. Instruction and curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment, 1*(2).

Bonniol, J. J. (1972). *Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires* [thèse de doctorat, Université de Provence].

Bonniol, J. J. (1973). Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires : Étude de quelques-uns de leurs déterminants. *Bulletin de Psychologie, 27*(309), 202-205. https://www.persee.fr/doc/bupsy_0007-4403_1973_num_27_309_10493

Borenstein, M., Hedges, L. V., Higgins, J. P. T. et Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Botelho, F., Madeira, R. A. et Rangel, M. A. (2015). Racial Discrimination in Grading: Evidence from Brazil. *American Economic Journal: Applied Economics, 7*(4), 37-52. https://doi.org/10.1257/app.20140352

Burgess, S. et Greaves, E. (2009). Test scores, subjective assessment and stereotyping of ethnic minorities. *CMPO Working Paper Series, 9*(221), 1-45. https://ifs.org.uk/publications/test-scores-subjective-assessment-and-stereotyping-ethnic-minorities

Burgess, S. et Greaves, E. (2013). Test scores, subjective assessment and stereotyping of ethnic minorities. *Journal of Labor Economics, 31*(3), 535-576. https://doi.org/10.1086/669340

Burnette, P. C. (2001). Elementary students' preferences for teacher praise. *The Journal of Classroom Interaction, 36*(1), 16-23. https://www.jstor.org/stable/23870540

Casanova, D. et Demeuse, M. (2011). Analyse des différentes facettes influant sur la fidélité d'une l'épreuve d'expression écrite d'un test standardisé de français langue étrangère. *Mesure et Évaluation en Éducation, 34* (1), 25-53.

Casanova, D. et Demeuse, M. (2017). Evaluateurs évalués : évaluation diagnostique des compétences en évaluation des correcteurs d'une épreuve d'expression écrite à forts enjeux. *Mesure et Évaluation en Éducation, 39* (3), 59-96.

Chouinard, R. (2002). Différences d'attitudes et de comportements en classe selon l'appartenance sexuelle. Dans J. Fijakow et T. Nault (dir.), *La gestion de classe* (p. 185-198). De Boeck.

Clancy, K., Bartolomeo, J.S., Richardson, D. et Wellford, C.F. (1981). Sentence Decision making: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity. *Journal of Criminal Law & Criminology, 72*, 524. http://doi.org/10.2307/1143005

Connolly, P., Keenan, C. et Urbanska, K. (2018). The trials of evidence-based practice in education: a systematic review of randomized controlled trials in education research 1980-2016. *Educational Research, 60*, 276-291. https://doi.org/10.1080/00131881.2018.1493353

Contreras, D. (2023) Gender differences in grading: teacher bias or student behaviour. *Education Economics*. https://doi.org/10.1080/09645292.2023.2252620

Crahay, M. (2019). *Peut-on lutter contre l'échec scolaire ?* (4e éd.). De Boeck supérieur.

Dachet, D. et Baye, A. (2021). Evidence-Based Education: The (Not So Simple) Case of French-speaking Belgium. *ECNU Review of Education*, *4*(1), 164-189 http://doi.org/10.1177/2096531120928086

Darley, J. M. et Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*(1), 0–33. https://doi.org/10.1037/0022-3514.44.1.20

Dee, T. S. (2004). Teachers, race and student acievement in a randomized experiment. *The Review of Economics and Statistics, 86*(1), 195-210. https://www.jstor.org/stable/pdf/3211667.pdf

Dee, T. S. (2005). A teacher like me: does race, ethnicity, or gender matter? *The American Economic Review, 95*(2), 158-165. http://doi.org/10.1257/000282805774670446

De Landsheere, G. (1980). *Évaluation continue et examens : précis de docimologie*. Labor.

Detheux-Jehin, M. et Kellens, M.-F. (1992). *APER secondaire – Recherche en vue de l'amélioration des pratiques d'évaluation - bilan et d'une réduction des redoublements*. Service de Pédagogie expérimentale de l'Université de Liège.

Di Liberto, A., Casula, L. et Pau, S. (2022). Grading practices, gender bias and educational outcomes: evidence from Italy. *Education Economics, 30*(5), 481-508. https://doi.org/10.1080/09645292.2021.2004999

Drouet, M., Duryea, S., Marotta, L., Rangel, M. A. et Guillén, L. R. (2023). *Discrimination in grading: evidence from immigrant in Ecuador*. Inter-American Development Bank.

Dubet, F. (2010). L'école « embarrassée » par la mixité. *Revue Française de Pédagogie, 171*, 77-86. https://doi.org/10.4000/rfp.1907

Fagnant, A., Duroisin, N. et Van Nieuwenhoven, C. (2014). L'évaluation au service des apprentissages. De l'évaluation formative en salle de classe au pilotage du système par les résultats des évaluations externes : l'évaluation peut-elle être au service des apprentissages scolaires ? *Education & Formation, e302*, 133-137.

Fefer, S., DeMagistris, J. et Shuttleton, C. (2016). Assessing adolescent praise and reward preferences for academic behavior. *Translational Issues in Psychologie Science, 2*(2), 153-162. https://doi.org/10.1037/tps0000072

Figlio, D. N. (2005). Names, expectations and the black-white test score gap. *Working Papers, 11195*, 1-32. http://doi.org/10.3386/w11195

Findlay, L. C. et Ste-Marie, D. M. (2004). A Reputation Bias in Figure Skating Judging. *Journal of Sport & Exercise Psychology, 26*(1), 154–166. http://doi.org/10.1123/jsep.26.1.154

Fresson, M. et Dardenne, B. (2019). Chapitre 4 – Le redoublants, victimes de stéréotypes dévalorisants ? Dans M. Crahay (dir.), *Peut-on lutter contre l'échec scolaire ?* (4e éd., p. 231-260). De Boeck supérieur.

Fresson, M., Dardenne, B. et Meulemans, T. (2018). Impact of diagnosis threat on neuropsychological assessment of people with acquired brain injury: Evidence of mediation by negative emotions. *Archives of Clinical Neuropsychology*. http://doi.org/10.1093/arclin/acy024

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research1. *Educational Researcher, 5*(10), 3-8. https://doi.org/10.3102/0013189X005010003

Grisay, A. (1984). *Trébucher sur le seuil de l'école. Enquête sur le problème du redoublement et de l'échec scolaire au premier cycle de l'enseignement primaire*. Laboratoire de Pédagogie expérimentale de l'Université de Liège.

Hanna, R. et Linden, L. (2009). Measuring discrimination in education. *Working Paper, 15057*, 1-34. http://doi.org/10.3386/w15057

Hinnerich, B. T., Höglin, E. et Johannesson, M. (2010). Are boys discriminated in Swedish high schools? *Economics of Education Review, 30*(4), 682-690. https://doi.org/10.1016/j.econedurev.2011.02.007

Hinnerich, B. T., Höglin, E. et Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Working Paper, 14*, 1-23. https://www.ifau.se/en/Research/Publications/Working-papers/2010/Are-boys-discriminated-in-Swedish-high-schools/

Hinnerich, B. T., Höglin, E. et Johannesson, M. (2015). Discrimination against students with foreign backgrounds: evidence from gradind in Swedish public high schools. *Education Economics, 23*(6), 660-676. http://doi.org/10.1080/09645292.2014.899562

Humphrey-Murto, S., Shaw, T., Touchie, C., Pugh, D., Cowley, L. et Wood, T. J. (2021). Are raters influenced by prior information about a learner? A review of assimilation and contrast effects in assessment. *Advances in Health Sciences Education, 26*, 1133-1156. http://doi.org/10.1007/s10459-021-10032-3

Jaegers, D., Lafontaine, D. et Dupont, V. (2023). Gendered aspirations to study math-intensive subjects: how perceived teacher expectations matter in French-speaking Belgium? Dans P. Watson, C. M. Rubie-Davies et B. Ertl (dir.), *The Routledge international handbook of gender, stereotype threat, and teacher expectations: Synergies and solutions*. Routledge.

Jussim, L. et Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Kahneman, D. Sibony, O. et Sunstein, C. R. (2021). *Noise. Pourquoi nous faisons des erreurs de jugement et comment les éviter*. Odile Jacob.

Lafontaine, D., Crahay, M. et Monseur, C. (2019). Chapitre 2 - Redoublement, efficacité et équité : l'éclairage des enquêtes internationales. Dans M. Crahay (dir.), *Peut-on lutter contre l'échec scolaire ?* (4e éd., p. 111-154). De Boeck supérieur.

Lafontaine, D. et Monseur, C. (2009). Les évaluations des performances en mathématiques sont-elles influencées par le sexe de l'élève ? *Mesure et Évaluation en Education, 32*(2), 1-106. https://doi.org/10.7202/1024955ar

Lafontaine, D. et Toczek-Capelle, M.-C. (2023). *L'évaluation en classe au service de l'apprentissage des élèves : rapport de synthèse*. Cnesco-Cnam.

Laugier, H. et Weinberg, D. (1936). *Commission française pour l'enquête Carnégie sur les examens et concours. La correction écrite des épreuves du baccalauréat*. Maison du Livre

Lavy, V. (2004). Do gender stereotypes reduce girls' human capital outcomes? Evidence from a natural experiment. *Working Paper, 10678*, 1-53. https://doi.org/10.3386/w10678

Leclercq, D. (1999). *Edumétrie et Docimologie*. Presses de l'Université de Liège.

Leclercq, D., Nicaise, J. et Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. Dans M. Demeuse (dir.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (p. 273-292). Les Editions de l'Université de Liège.

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J. et Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: a systematic review. *Applied Measurement in Education, 33*(2), 124-140. https://doi.org/10.1080/02602938.2018.1545896

Leyens, J.-P. et Yzerbyt, V. (2017). *Psychologie sociale : un outil de référence.* Mardaga.

Lindahl, E. (2007). Comparing teachers' assessments and national test results – evidence from Sweden. *Working Paper, 24*, 1-20. https://www.ifau.se/en/Research/Publications/Working-papers/2007/Comparing-teachers-assessments-and-national-test-results---evidence-from-Sweden/

Lord, C. G., Ross, L. et Lepper, M. R. (1979). Biaised assimilation and attitude polarization: the effects of prior theories on subsequently considerd evidence. *Journal of Personality and Social Psychology, 37*(11), 2098-2109. https://doi.org/10.1037/0022-3514.37.11.2098

Malouff, J. M., Emmerton, A. J. et Schutte, N. S. (2013). The risk of a halo bias as a reason to keep students anonymous during grading. *Teaching Psychology, 40*(3), 233-237. https://doi.org/10.1177/0098628313487425

Malouff, J. M. et Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education, 60*(3), 245-256. https://doi.org/10.1177/0004944116664618

Martin, J. (2002). Aux origines de la « science des examens » (1920-1940). *Histoire de l'Éducation, 94*, 177-199. https://doi.org/10.4000/histoire-education.817

Martinot, P., Colnet, B., Breda, T., Sultan, J., Touitou, L., Huguet, P., Spelke, E., Dehaene-Lambertz, G., Bressoux, P. et Dehaene, S. (2025). Rapid emergence of a maths gender gap in first grade. *Nature, 643*, 1020–1029. https://doi.org/10.1038/s41586-025-09126-4

Mechtenberg, L. (2009). Cheap talk in the classroom: how biased gradind at school explains gender differences in achievements, career choices and wages. *Review of Economic Studies, 76*, 1431-1459. http://doi.org/10.1111.j.1467-937X.2009.00551.x

Monseur, C. et Baye, A. (2015). *Les inégalités scolaires d'origines sociale et ethno-culturelle : une possible amplification*. Conseil National d'Évaluation du Système Scolaire. http://hdl.handle.net/2268/193830

Nieminen, L. R. G., Rudolph, C. W., Baltes, B. B., Casper, C. M., Wynne, K. T. et Kirby, L. C. (2013). The combined effect of ratee's bodyweight and past performance information on performance judgments. *Journal of Applied Social Psychology, 43*, 527-543. http://doi.org/10.1111/j.1559-1816.2013.01033.x

Ouazad, A. (2008). Assessed by a teacher like me: race, gender and subjective evaluations. *INSEAD Working Paper, 57*, 1-44. http://doi.org/10.2139/ssrn.1267109

Palmer, J. K. et Gore, J. S. (2014). A theory of contrast effects in performance appraisal and social cognitive judgments. *Psychological Studies, 59*(4), 323-336.

Pansu, P. et Bressoux, P. (2003). *Quand les enseignants jugent leurs élèves*. Presses Universitaires de France.

Piéron, H. (1963). *Examens et docimologie.* Presses universitaires de France.

Protivinsky, T. et Munich, D. (2018). Gender bias in teahcers' grading: whats is in the grade. *Studies in Educational Evaluation, 59*, 141-149. http://doi.org/10.1016/j.stueduc.2018.07.006

Rangvid, B. S. (2015). Systematic differences across evaluation schemes and educational choice. *Economics of Education Review, 48*, 41–55. http://doi.org/10.1016/j.econedurev.2015.05.003

Rauw, K., Dachet, D. et Baye, A. (2023). *Les performances des élèves redoublants sont-elles sous-estimées* ? [Communication orale]. 34e colloque de l'ADMEE-Europe – « évaluer les apprentissages : continuités et ruptures », Mons, Belgique. https://hdl.handle.net/2268/301504

Raynal, F. et Rieunier, A. (2014). *Pédagogie, dictionnaire des concepts clés : apprentissage, formation, psychologie cognitive* (10e éd.). ESF éditeur.

Romainville, M. (2023). *A l'école du doute : apprendre à penser juste en découvrant pourquoi on pense faux*. Presses Universitaires de France

Rosenthal, R. A. et Jacobson, L. (1971). *Pygmalion à l'école*. Casterman.

Rot, N. et Bujas, Z. (1959). Les distributions des notes scolaires comparées aux distributions des résultats obtenus aux tests de connaissances. *Le Travail Humain, 22*(1-2), 19-26. https://www.jstor.org/stable/40659144

Rouse, K. et Hunziker, B. (2020). Child bodyweight and human capital: Test scores, teacher assessments and noncognitive skills. *Economics of Education Review, 79*(102042), 1-19. https://doi.org/10.1016/j.econedurev.2020.102042

Royer, D. J., Lane, K. L., Dunlap, J. D. et Parks Ennis, R. (2018). A systematic review of teacher-delivered behavior-specific praise on k-12 student performance. *Remedial and Special Education, 40*(2), http://dx.doi.org/10.1177/0741932517751054

Sanrey, C., Bressoux, P., Lima, L. et Pansu, P. (2021). A new method for studying the halo effect in teacher's judgement and its antecedents: bringing out the role of certainty. *British Journal of Educational Psychology, 91*, 658-675. http://doi.org/10.1111/bjep.12385

Shaw, T., Wood, T. J., Touchie, C., Pugh, D. et Humphrey-Murto, S. M. (2021). How biased are you? The effect of prior performance information on attending physician ratings and implication for learner handover. *Advances in Health Sciences Education, 26*, 199-214. http://doi.org/10.1007/s10459-020-09979-6

Slavin, R. E. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher, 31*(7), 15-21. https://doi.org/10.3102/0013189X031007015

Slavin, R. E. (2007). *Educational Research in an Age of Accountability*. Pearson Education.

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics, 45*, 523–538. http://doi.org/10.1007/s00181-012-0609-x

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M. et Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811. https://doi.org/10.1037/0022-3514.69.5.797

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C. et Sharot, T. (2017). How People Update Beliefs about Climate Change: Good News and Bad News. *Cornell Law Review, 102*(6), 1431-1443. https://scholarship.law.cornell.edu/clr/vol102/iss6/1/

Tenenbaum, H. R. et Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253–273. https://doi.org/10.1037/0022-0663.99.2.253

Tversky, A. et Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Van Ewijk, R. (2011). Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review, 30*, 1045-1058. https://doi.org/10.1016/j.econedurev.2011.05.008

Zavodny, M. (2013). Does weight affect children's test scores and teacher assessments differently? *Economics of Education Review, 34*, 135-145. https://doi.org/10.1016/j.econedurev.2013.02.003

Chapitre 10. La mesure dans la langue de Molière

Etat des lieux et avenues à envisager

Dominique Casanova et Sébastien Béland

I. Introduction

Good measurement leads to profound theory.

John « Mike » Linacre

Dans un article écrit à l'occasion des 30 ans de la revue *Mesure et évaluation en éducation,* Pierre Valois et Romain Martin s'inquiétaient du manque de publications dans cette revue portant sur les modèles de mesure (de caractéristiques individuelles : connaissances, acquis de l'expérience…), en comparaison d'autres revues anglo-saxonnes. Ces publications s'inscrivant majoritairement dans le cadre de la théorie classique des tests et de la théorie de la généralisabilité, ils concluaient leur article de la manière suivante :

> Il apparaît évident que la revue Mesure et évaluation en éducation doit être plus active sinon dans le développement théorique et mathématique des modèles de réponses aux items (MRI), qui relève davantage des experts en (…) [mesure], du moins dans l'application de ces modèles au monde de l'éducation. Depuis trente ans, la publication d'articles empiriques sur les MRI est insuffisante compte tenu du potentiel évident de ces modèles pour valider des outils d'évaluation et estimer les

> processus cognitifs et affectifs sous-jacents aux patrons de réponse individuels. (Valois et Martin, 2008, p. 149-150).

Dix-huit ans plus tard, pour les 40 ans de l'ADMEE-Europe, nous avons souhaité dresser un nouvel état des lieux en la matière. Nous constatons une part croissante de communications portant sur le thème de la mesure. Si la pratique de la mesure reste très ancrée dans la théorie classique des tests, davantage d'articles empiriques portant sur les MRI sont publiés, témoignant d'une plus grande appropriation de ces modèles. Toutefois, les articles proposés informent peu sur les évolutions récentes dans le domaine de la mesure.

C'est pourquoi, dans la seconde partie de ce chapitre, nous présentons un florilège des avancées dans ce domaine. Nous espérons qu'elles éveilleront la curiosité du public de la revue et amèneront les lecteurs à creuser des pistes innovantes et à partager leurs découvertes avec le reste de la communauté au sein de la revue *Mesure et évaluation en éducation* de l'ADMEE.

II. Qu'est-ce que la mesure ?

Selon l'Ordre des psychologues du Québec, la psychométrie « se définit comme la mesure des processus mentaux et des attributs propres aux individus ». Ces phénomènes peuvent être observés de différentes manières, notamment par le remplissage d'une grille d'observation ou par la collecte d'une série de réponses à un questionnaire. Un tel questionnaire peut être constitué par exemple d'affirmations pour lesquelles l'individu observé précise un degré d'accord avec l'énoncé, ou d'exercices qu'il est invité à résoudre en sélectionnant ou en communiquant la réponse appropriée.

Ces instruments de mesure (grille, questionnaire…) sont conçus de sorte à représenter un construit théorique : le ou les fameux attributs latents que la chercheuse ou le praticien souhaite mettre en évidence (le concept objet de la mesure peut recouvrir plusieurs dimensions sous-jacentes). Ici, il existe deux grandes conceptions philosophiques de la relation entre le trait latent (ce que nous tentons de mesurer et qui est non observable, par exemple, la motivation scolaire ou la compréhension en algèbre) et les réponses données aux items par une personne évaluée. Par convention,

nous représentons le trait latent par un cercle et les questions à un test par des carrés (voir la Figure 1).

Le modèle réflectif (à gauche) comme celui des MRI suppose que le trait latent cause les réponses aux items. Cette approche s'inscrit dans une perspective épistémologique réaliste, selon laquelle le trait existe indépendamment de l'individu observant et peut être inféré à partir des

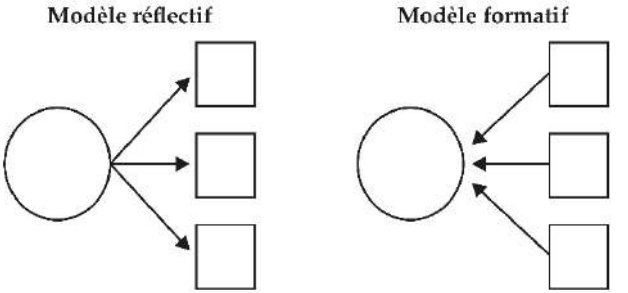


Figure 1 : Représentation d'un modèle formatif et d'un modèle réflectif

comportements observés. Le modèle formatif à droite, comme le *log-multiplicative association* (Anderson et Yu, 2007), adopte une perspective constructiviste : ici, ce sont les réponses aux items qui définissent le trait latent. Le trait n'est pas une cause mais une construction empirique fondée sur les indicateurs observés.

Une attention particulière doit être portée à l'élaboration des instruments de mesure, en s'appuyant sur l'état de la recherche en la matière et l'avis de personnes expertes. L'instrument est ensuite expérimenté, révisé puis mis à l'épreuve afin de recueillir des éléments de preuve de la pertinence de son utilisation dans un contexte donné (dimensionnalité, fidélité, fonctionnement différentiel des items…).

Selon le contexte, les personnes concevant l'instrument de mesure s'intéressent aux résultats de groupes d'individus ou aux résultats individuels. Elles peuvent vouloir comparer les résultats d'individus entre eux, à ceux d'un échantillon de référence ou à des valeurs-seuils pour procéder à l'interprétation des résultats. Entre alors inévitablement en jeu la question des erreurs de mesure liées à l'utilisation de l'instrument.

Il est fondamental de comprendre que les modèles de mesure sont teintés d'une humilité liée à leur forme : l'erreur de mesure est inévitable. Elle a évidemment de multiples sources, mais mentionnons les limites de l'instrumentation (un instrument ne permet de recueillir qu'un nombre limité d'informations, à partir de formulations qui ne reflètent pas parfaitement le construit objet de la mesure), les conditions matérielles dans lesquelles la mesure est réalisée (qui peuvent varier d'une occasion à une autre, en dépit des recommandations) et les dispositions dans lesquelles se trouvent les individus lors de la collecte d'information (motivation, fatigue…).

La question de l'erreur de mesure est un élément différenciant des trois principales théories de mesure utilisées dans le monde de l'éducation. Dans la théorie classique des tests, l'erreur de mesure est établie globalement, à partir d'une mesure de la fidélité de l'instrument, et s'applique de manière uniforme tout au long de l'échelle. Un des grands avantages de la théorie de la généralisabilité est de distinguer différentes facettes constituant le dispositif de mesure et d'identifier la part d'erreur de mesure qui peut être attribuée à chacune des facettes d'instrumentation et à leurs interactions. La théorie de la réponse aux items, pour sa part, permet d'estimer l'erreur de mesure de l'instrument à chaque emplacement sur l'échelle de mesure. Une comparaison empirique de modèles relevant de ces trois théories est proposée par Bertrand (2003). Enfin, mentionnons que lorsqu'une chercheuse ou un praticien s'intéresse à la fidélité des scores, elle ou il voudra estimer la capacité d'un instrument à générer de l'erreur.

III. Le contenu de la revue Mesure et évaluation en éducation : le cas de la mesure

Dans notre synthèse des publications de *Mesure et évaluation en éducation*, nous avons pris en considération tous les articles qui concernent l'évaluation d'un trait latent. Nous avons inclus tous les articles qui abordent la mesure via des présentations de modèles ou de théories de mesure, des considérations sur les conditions d'application de ces modèles, des mises à l'épreuve d'instruments de mesure, des analyses de l'effet de traitements sur l'évolution du trait latent via deux mesures successives, etc.

Quoique fort intéressantes pour le domaine de l'éducation, nous n'avons pas tenu compte des communications qui portent sur la part qui peut être

attribuée à des facteurs externes (effet classe, effet établissement…) dans la variance des résultats des individus. Le public intéressé par ces pratiques pourra se référer, entre autres, à Bressoux (1994), Bressoux et al. (1998), Demeuse et Monseur (1999), Le-Bastard-Landrier et Suchaut (2000) et Dumay et Dupriez (2005). Nous n'avons pas inclus non plus les recensions d'ouvrages, quand bien même ces ouvrages portaient sur la mesure, ni les traductions anglaises d'articles parus préalablement en français, ni les pages introductives de certains numéros de la revue.

III.a. Les thèmes étudiés dans la revue MEE

Nous avons identifié 179 articles traitant de mesure (sur un ensemble de 638 articles, soit 28% du total). Un premier constat est donc que, quoiqu'il s'agisse du premier mot du titre de la revue *Mesure et évaluation en éducation*, la mesure ne représente qu'une part limitée des publications. Sa présence a cependant progressé dans le temps, notamment depuis la fin des années 2000 (Cf. Figure 2), puisqu'il y a aujourd'hui de l'ordre de 5 à 6 articles par an se rapportant à la mesure, soit près de deux articles par numéro en moyenne.

.

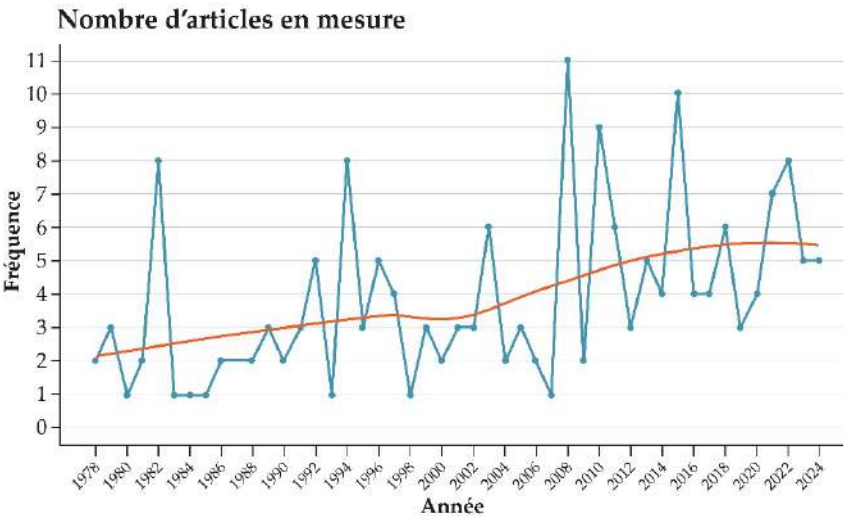


Figure 2 : Évolution du nombre d'articles portant sur la mesure dans la revue Mesure et évaluation depuis son origine

Nous avons ensuite considéré l'ensemble des titres des articles, ainsi que l'ensemble des résumés, pour mener des analyses textuelles automatisées, de sorte à rendre compte des thèmes abordés dans ces articles et de leur progression au fil du temps. La Figure 3 présente un nuage de mots qui émerge de l'ensemble des titres des articles. Nous n'avons conservé que les lemmes qui se rapportaient à la mesure, en ne comptabilisant chaque lemme qu'une fois au sein de chaque titre. La taille de police de chacun des lemmes représentés rend compte de sa fréquence relative d'apparition dans le corpus et nous nous sommes restreints aux lemmes apparaissant au minimum 7 fois. Le lemme *mesure* est le plus présent : il apparaît 35 fois (et le lemme *mesurer* apparaît 7 fois).

Figure 3 : Nuage des lemmes les plus fréquents des titres d'articles de la revue Mesure et évaluation depuis son origine

Nous pouvons résumer ce nuage de mots comme suit : les articles retenus portent avant tout sur :

- la mesure ;
- principalement sur la validation ou la validité (psychométrique) ;
- d'une échelle, d'un instrument, d'un test ou d'un questionnaire ;
- élaboré(e), adapté(e) ou développé(e) ;
- à destination d'étudiant-e-s, d'enseignant-e-s ou d'élèves ;
- dans un contexte de santé, science, éducation ou scolaire ;
- en s'appuyant sur des modèles, des théories ou des approches.

Les titres d'articles sont toutefois inévitablement succincts et peuvent ne refléter que partiellement le contenu exposé. Un travail similaire a donc été entrepris sur le contenu des résumés des articles. Il est toutefois limité, pour des raisons techniques, aux articles postérieurs à 2007, date à partir de laquelle nous disposons de leur version numérique. Si ce nuage comporte beaucoup de lemmes communs à ceux résultant de l'analyse des

titres, la fréquence de certains termes se renforce et de nouveaux lemmes apparaissent (Figure 4). Les termes *item*, *psychométrique* et *qualité*, par exemple, prennent une place plus importante. Nous voyons également apparaître des termes comme *factoriel*, *confirmatoire*, *exploratoire*, *facteur*, *généralisabilité* et *fidélité*. Cela nous renseigne sur les modèles et les techniques les plus fréquemment mobilisés.

Figure 4 : Nuage des mots les plus fréquents des résumés d'articles de la revue Mesure et évaluation depuis 2004

Toutefois, une telle représentation ne permet pas de voir l'évolution des tendances dans le temps. Il convient pour cela d'entrer un peu plus dans le contenu des articles. Nous décrivons ci-dessous ce que nous avons perçu comme tendances d'évolution concernant la réflexion sur la mesure et sa mise en œuvre au sein de la revue *Mesure et évaluation en éducation.*

1978 - 1984

Durant les premières années de la revue, alors que l'ADMEE n'était que canadienne, le centre d'intérêt principal était la mesure de la progression des apprentissages, par des procédures de test-retest (voir par exemple Larose et Racine, 1981). Les concepts de fidélité, validité, d'analyse d'items et de scores sont évoqués dès 1979 (Racine, 1979) et les MRI sont abordées une première fois en 1982 (Louis, 1982). C'est une période où la revue se structure, passant progressivement de comptes-rendus d'ateliers et de conférences à des articles dont la forme correspond aux standards scientifiques actuels.

1985 - 1994

À partir de 1986 apparaissent des communications régulières sur les MRI (Séguin et Augier, 1986 ; Bertrand, 1988) et sur la généralisabilité (notamment à l'occasion d'un numéro spécial dédié en 1989 à la mémoire d'Yvan Tourneur). Un autre numéro spécial est consacré à la mesure en 1994, qui regroupe une version française des actes des deuxième et troisième *International Symposium on Educational Testing*.

1995 – 2004

Il s'agit de la décennie qui compte le plus d'occurrences du terme *mesure*, avec notamment la publication en 2003 d'un numéro spécial sur la généralisabilité, et du terme *instrument*, avec la structuration d'une démarche de développement/traduction/adaptation et de validation d'instruments de mesure, ancrée dans la théorie classique des tests autour de l'analyse des coefficients de discrimination, de la fidélité interne, la mise en œuvre d'analyses factorielles et la corrélation avec un/des critère(s) externe(s) (voir par exemple Louis et Trahan, 1995 ; Lafontaine, 1999).

2005 – 2014

Les entreprises d'élaboration et de validation (terme qui s'impose durant cette période) se poursuivent, et le terme *échelle* supplante les termes *instrument*, *questionnaire* et *test*. Un nouveau numéro spécial sur la mesure, regroupant les actes d'une rencontre organisée par l'Institut de recherche et de documentation pédagogique (IRDP), paraît en 2008. Il est de plus en plus fréquemment fait référence au modèle de Rasch, mais de nouveaux modèles de mesure font également leur apparition : les modèles de classification diagnostique (Loye, 2005 et 2010). Un nombre croissant de

communications portent un regard plus acéré sur l'utilisation des modèles de mesure et l'estimation des paramètres (selon les logiciels utilisés ou en éliminant l'influence d'individus présentant des patrons de réponses atypiques). L'utilisation parfois abusive de certains indices, comme l'alpha de Cronbach, est également soulignée (Laveault, 2012).

2015 – 2024

Les communications portant sur l'élaboration ou la validation d'échelles sont désormais majoritaires parmi les communications portant sur la mesure. Les termes *validation* (ou *validité*) et *échelle* sont installés, de même le terme *élaboration* supplante le terme *développement*. Les études menées s'intéressent davantage aux étudiants et les domaines de sciences et de la santé font une apparition marquée. Pour la première fois dans la revue, un article propose une analyse empirique à l'aide d'un modèle de classification diagnostique (Duong Thi et Loye, 2019). À signaler également, un travail théorique que nous livre Laurencelle (2015, 2016 et 2017) sur une théorie des seuils psychométriques à double contrôle d'erreur.

Synthèse

Un des enseignements de cette revue chronologique est la part croissante des articles qui portent sur l'élaboration (ou adaptation ou traduction) d'une échelle (ou instrument ou questionnaire) ou sa validation au sein de la revue (voir le Tableau 1).

Période	Article sur l'élaboration/validation d'une échelle	Total articles	Pourcentage
1978 - 1984	1	18	6%
1985 - 1994	7	30	23%
1995 - 2004	13	32	41%
2005 - 2014	15	43	35%
2015 - 2024	34	56	61%

Tableau 1 : Part des articles portant sur l'élaboration ou la validation d'une échelle par décennie

Si ces articles constituent une véritable richesse pour la communauté et ouvrent la porte à des échanges entre pairs ainsi qu'à des méta-analyses, ils masquent la stagnation du nombre de communications portant sur les modèles de mesure et leurs conditions d'application. Or, les choses évoluent extrêmement rapidement dans ce domaine, comme en témoigne la section suivante.

IV. Quoi de neuf en mesure ?

Nous allons maintenant aborder quelques récents développements en mesure qui nous semblent prometteurs. Ce qui suit n'est évidemment pas exhaustif et doit être considéré comme une sélection subjective. Toutefois, cela permettra au lecteur de constater certaines tendances fascinantes qui se dégagent de plusieurs chantiers en mesure.

Avant d'aller plus loin, nous allons mentionner deux sujets qui semblent en perte de vitesse. Rappelons que Kuhn (1962) montrait que l'intérêt pour certaines théories ou approches peut varier dans le temps, selon les intérêts changeant d'une discipline. En effet, les intérêts d'une communauté scientifique sont historiquement situés : ce qui est jugé important n'est jamais éternel (Fleck, 1979)

D'abord, le nombre d'études publiées sur les statistiques de détection des réponses inappropriées (nommée *person-fit*) a un peu ralenti. En guise d'exemple, Snijders (2001) avait proposé une stratégie pour corriger la distribution de l'indice *lz*, nommée *lz**, qui avait été étendue aux items ordinaux par Sinharay (2016). Depuis, les travaux touchent davantage à d'autres éléments concernant les réponses inappropriées, par exemple la

détection de réponses aléatoires par des robots dans les questionnaires en ligne (Falk et al., 2025).

Ensuite, les développements sur la fidélité des scores ont été très importants depuis 20 ans, mais nous constatons un ralentissement dans le développement des coefficients et du nombre d'articles qui ont, notamment, critiqué le très connu alpha de Cronbach. Nous mentionnons toutefois des études de simulation portant sur le comportement des coefficients sur des items à réponse ordinale. Les travaux de Béland et Falk (2022) ainsi que Cho et Béland (2025) ont montré que le coefficient alpha est légèrement moins efficace que d'autres coefficients. De plus, les coefficients conçus pour les réponses continues, ou fondés sur des modèles comme l'analyse factorielle ou les MRI, se révèlent plus efficaces que ceux destinés aux réponses discrètes (ce qui concorde avec Robitzsch, 2020). Nous observons que les développements les plus intéressants sur la fidélité touchent à l'adaptation à différents modèles, dont les modèles multiniveau (Lin et al., 2025), qui peuvent modéliser des plus petits groupes tels que des classes dans une école.

IV.a. Une explosion des travaux mobilisant l'intelligence artificielle (IA)

Le programme des colloques des groupements qui s'intéressent à la mesure tels que l'*International meeting of the psychometric society* (IMPS) et de l'*International testing commission* (ITC) ont vu leurs propositions en IA augmenter substantiellement depuis la fin de la pandémie.

Les premiers travaux marquants se sont concentrés sur l'automatisation de la correction au test (Chen et al., 2010 ; Hussein et al., 2019) d'alléger le fardeau associé à cette tâche pour les personnes chargées de la correction. Le domaine a tellement généré de travaux que des revues systématiques sont disponibles ; telles que celle de Ramesh et Sanampudi (2022) qui expliquait que les systèmes actuels fonctionnent surtout pour les questions à choix multiples, mais que des développements prometteurs commencent à se faire pour les réponses courtes. Ils déclaraient qu'évaluer une dissertation en tenant compte de tous les paramètres tels que la pertinence du contenu par rapport à la consigne, le développement des idées, la cohésion et la cohérence demeure un important défi technique.

Les choses ont évolué tellement rapidement que des défis jugés difficilement surmontables dans cette revue systématique sont maintenant dépassés. Par exemple, le package en téléchargement libre `R-text` (Kjell

et al., 2023) permet de faire des analyses de textes très utiles en éducation. Dans bien des établissements scolaires, une enseignante pourrait souhaiter remplacer la correction manuelle de réponses écrites par une méthode plus rapide. En utilisant la fonction `textEmbed` de ce package, elle transforme les réponses textuelles des participants en vecteurs sémantiques, avant d'entraîner un modèle prédictif pour prédire les scores de l'épreuve analysée. Cela est utile, car la correctrice peut ensuite facilement prédire les scores pour les réponses de nouveaux participants. Bien qu'imparfaite, cette approche permet ainsi une notation automatisée tout en tenant compte de la richesse linguistique des réponses.

Les chercheurs travaillent aussi activement à optimiser ces procédures d'analyse. Par exemple, Jung (2025) proposait de faire une compression des données avant d'automatiser la correction. Il est intéressant de savoir que les algorithmes d'IA montrent des résultats prometteurs pour aider les personnes chargées de l'évaluation (voir le cas de Berkovsky et al., 2019). Nous mentionnons l'étude de Altemeyer (2019), où il était montré que l'IA peut réduire les biais humains lors de la mesure.

Enfin, les enquêtes à grande échelle comme le PISA et le TIMSS offrent des défis importants d'estimation des modèles de mesure utilisés traditionnellement. Nous observons de nombreux travaux en cours visant à améliorer l'estimation de l'habileté des élèves à l'aide de l'IA lors de ces grandes enquêtes (Ali, 2025 ; Ren, 2025).

IV.b. Les modèles de réseaux pour mieux comprendre les relations de dépendance entre les items de vos épreuves

Epskamp et al. (2018) montrent que ces modèles permettent non seulement de nouvelles interprétations des modèles de mesure traditionnels, mais ouvrent aussi la voie à des analyses conceptuelles et méthodologiques enrichies, notamment par l'utilisation des champs aléatoires de Markov pour représenter les relations de dépendance entre les items. Par exemple, une extension fondée sur les modélisations par équations structurelles (Epskamp, 2021) a été développée et rendue disponible dans le logiciel en accès libre R. Des indices classiques d'adéquation, tels que le RMSEA, le CFI et le TLI, sont aussi adaptés pour ce type de modèles (Golino et al., 2021).

Des exemples d'applications des modèles de réseau ont aussi été développés pour estimer le nombre de dimensions dans les items. Ces alternatives aux approches heuristiques du test de l'éboulis de Cattell et de la règle de Kaiser sont nichées dans l'analyse exploratoire de graphe (*exploratory graph analysis,* Golino et Epskamp (2017)). Elles permettent de visuellement représenter quels items se regroupent pour former une dimension telle que dans l'exemple de la Figure 5.

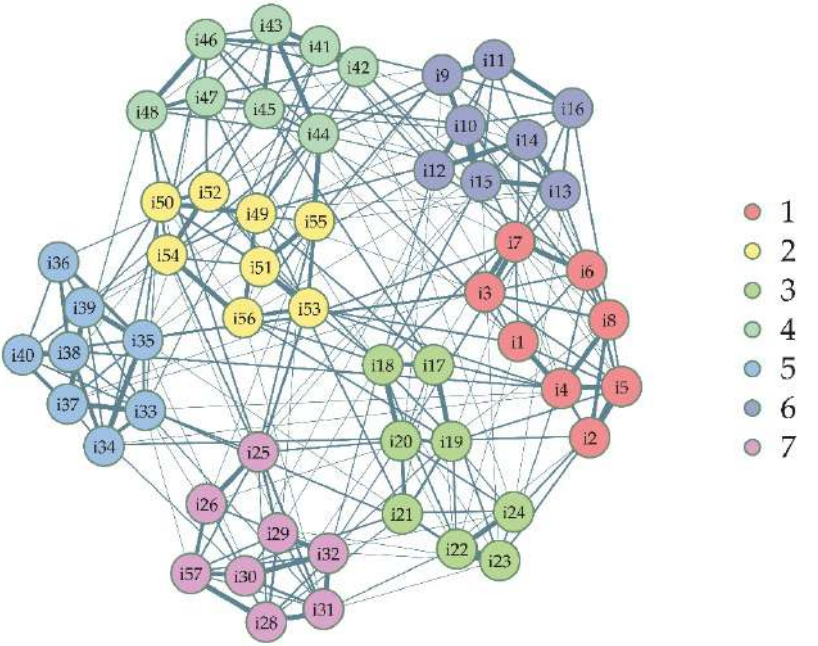


Figure 5 : Figure présentée par Golino et Epskamp (2017, p. 22) montrant sept dimensions latentes dans les données du Test de développement du raisonnement inductif. Des modèles cognitifs qui sont toujours aussi instructifs

Les modèles cognitifs de classification diagnostique visent à identifier les forces et faiblesses spécifiques des individus en lien avec des compétences latentes plutôt qu'un simple score global de performance. Ces modèles reposent sur une matrice Q, qui indique quelles compétences sont requises pour répondre correctement à chaque item d'un test. Plus précisément, chaque ligne de la matrice Q représente un item et chaque colonne une compétence ; une valeur de « 1 » signifie que l'attribut est requis pour réussir l'item, et « 0 » qu'il ne l'est pas. Ainsi, ils permettent une analyse plus fine des profils d'apprentissage et sont utiles en éducation pour fournir un diagnostic personnalisé des besoins des apprenantes et des apprenants.

Les développements récents visent à étendre leur possibilité d'application à un large spectre de situations. Par exemple, pour voir comment ils classifient les personnes répondantes en testing adaptatif (Bediwy, 2025)

ou pour détecter le fonctionnement différentiel d'items (Zhili, 2025). Leur potentiel est immense pour mieux comprendre les processus cognitifs des élèves et nous partageons l'avis de Madison (2025) qu'il est (fort) étonnant qu'ils ne soient pas davantage mobilisés dans les enquêtes à grande échelle et dans les écoles européennes.

IV.c. La validité : vers de nouveaux standards en 2027 !

Les écrits de Meehl (voir Grove, 2005) et de Borsboom (par exemple, son article de 2006) ont profondément influencé la réflexion sur la mesure en éducation et en psychologie. Ils ont notamment contribué à poser des questions fondamentales sur la nature des tests psychométriques. Par exemple, la relation entre l'intelligence et le score de QI soulève une interrogation centrale : est-ce que le test représente fidèlement l'intelligence, ou est-ce qu'il la cause (voir la Figure 1 qui illustre ce questionnement) ? Ces deux perspectives mènent évidemment à des interprétations très différentes.

Pour y voir plus clair, il faut s'interroger sur ce qui se passe réellement dans l'esprit d'une personne lorsqu'elle passe un test de QI : quelles habiletés mobilise-t-elle, quelles stratégies adopte-t-elle, et dans quel contexte s'inscrit sa performance ? Il est également crucial de comprendre ce que le test lui-même fait : quelles tâches il propose, quelles réponses il attend, et selon quelles règles il évalue ces réponses. À partir de tout cela, un score est produit, un chiffre souvent perçu comme objectif, mais qui résulte en réalité d'une série de choix méthodologiques, statistiques et théoriques.

À en juger par le nombre croissant de publications sur le sujet, la question de la validité demeure un thème central depuis une quinzaine d'années. Et cet intérêt ne risque pas de faiblir, puisque la version actuelle des *Standards for Educational and Psychological Testing* est en cours de révision. Des consultations publiques sont prévues entre 2025 et 2027. Rappelons que la dernière version remonte à 2014, et que des discussions majeures sont en cours sur l'impact des tests en matière d'équité et d'égalité, ainsi que sur la validité liée aux processus de réponse.

IV.d. Les modélisations bayésiennes se démocratisent (enfin)

Les modélisations bayésiennes sont de plus en plus faciles à utiliser et les développements dans ce paradigme statistique sont toujours importants. Par exemple, une version bayésienne du package R `lavaan`, nommé `blavaan` (Merkle, 2025), est maintenant disponible pour faire des analyses factorielles et des modélisations par équations structurelles bayésiennes avec du code très simple.

La personne intéressée par la mesure constatera que ces modèles sont très importants et offrent une alternative pertinente en présence d'échantillons de petite taille (Swaminathan et Gifford, 1982 ; van de Schoot et Miocevic, 2020). Fox (2010) rapportait d'ailleurs que l'estimation dans ces modèles, et en présence d'une bonne information *a priori*, génère moins d'erreurs types que l'estimateur fréquentiste par maximum de vraisemblance.

IV.e. Les modèles de mesure modernes sont toujours en développement

Les MRI traditionnels permettent d'évaluer l'effet d'un traitement à partir de réponses, supposant que tous les items d'un test réagissent de la même façon. L'approche IL-HTE (*item-level heterogeneous treatment effects*) développée par Gilbert (2024) va plus loin en examinant l'effet du traitement pour chaque item individuellement. Cela permet de repérer des effets ciblés : certains items montrent une amélioration marquée, d'autres ne témoignent d'aucun changement, voire d'un effet négatif. Il devient donc possible de révéler des effets cachés qui seraient invisibles avec une analyse globale traditionnelle.

Il y a aussi des développements sur d'autres modèles pertinents, en éducation, qui étaient moins populaires ces dernières années. Un exemple est la Théorie de la réponse à l'item séquentielle inspirée de Tutz (1997). Contrairement aux modèles habituels qui se concentrent sur les réponses correctes ou incorrectes à un test donné, ce modèle s'intéresse aux processus de réponse sur plusieurs tentatives successives d'un même test, en particulier lorsqu'un test a été initialement échoué. De récents développements pour estimer les paramètres de ce modèle sont très prometteurs (Lu, 2025).

V. Conclusion

L'analyse des publications de la revue *Mesure et évaluation en éducation* montre que la part des articles portant sur la mesure est en progression. Toutefois, une majorité des articles de mesure s'intéressent à l'élaboration et à la validation d'échelles. La réflexion théorique ou les retours d'expérience sur les modèles de mesure existants, leurs conditions d'application et l'émergence de nouveaux modèles restent limités.

Or, les avancées dans le domaine progressent à un rythme sans précédent et l'IA y joue désormais un rôle central. Les modèles sont devenus plus souples que jamais, capables de traiter une grande variété de réponses et de situations, qu'il s'agisse de QCM, de textes ou du nombre de passations. On peut même modéliser le contexte social de mesure (Mislevy, 2018). Ils offrent ainsi un soutien plus solide aux arguments de validité que par le passé. Les modèles diagnostiques, en particulier, permettent de formuler des inférences plus précises sur les apprentissages des élèves, en identifiant finement leurs forces et leurs besoins.

Enfin, si tout se déroule comme prévu, de nouveaux standards devraient voir le jour d'ici 2027, avec une attention particulière accordée à la question de l'impact des tests en matière d'équité et d'égalité. Nous espérons que les articles de *Mesure et évaluation en éducation* reflèteront ces évolutions !

VI. Bibliographie

Ali, U. (2025, 16 juillet). A two-step imputation approach combining IRT and deep-learning methods in large-scale survey assessments [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Altemeyer, B. (2019, 16 avril). Making the business case for AI in HR: two case studies. Strategic HR Review 18(2), 66–70. https://doi.org/10.1108/SHR-12-2018-0101

Bediwy, A. (2025, 16 juillet). Evaluating the classification accuracy of an adaptive diagnostic test [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Béland, S. et Falk, C. F. (2022). A comparison of modern and popular approaches to calculating reliability for dichotomously scored items. Applied Psychological Measurement. https://doi.org/10.1177/01466216221084210

Bertrand, R. (1988). Pourquoi de nouvelles théories de la mesure ? Mesure et évaluation en éducation, 11(2), 5-25.

Bertrand, R. (2003). Une comparaison empirique de modèles de la théorie classique, de la théorie de la généralisabilité et de la théorie des réponses aux items. Mesure et évaluation en éducation, 26(1-2), 75–89. https://doi.org/10.7202/1088241ar

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J. et Kleitman, S. (2019, 4–9 mai). Detecting personality traits using eye-tracking data [communication orale]. CHI Conference on Human Factors in Computing Systems, Écosse, Royaume-Uni. https://doi.org/10.1145/3290605.3300472

Borsboom, D. (2006). The attack of the psychometricians. Psychometrika, 71(3), 425–440. https://doi.org/10.1007/s11336-006-1447-6

Bressoux, P. (1994). Estimer et expliquer les effets des classes : les cas des acquisitions en lecture. Mesure et évaluation en éducation, 17(1), 75–94.

Bressoux, P., Leroy-Audouin, Ch. et Coustère, P. (1998) Les extensions des modèles multiniveau et leur application pour l'évaluation en éducation. Mesure et évaluation en éducation, 21(1), 35–59. https://doi.org/10.7202/1091357ar

Chen Y.-Y., Liu C.-L., Chang T.-H. et Lee C.-H. (2010). An unsupervised automated essay scoring system. IEEE Intelligent Systems 25, 61–67. https://doi.org/10.1109/MIS.2010.3

Cho, E. et Béland, S. (2025). Reliability in unidimensional ordinal data: A comparison of continuous and ordinal estimators. Psychological Methods. https://doi.org/10.1037/met0000739

Demeuse, M. et Monseur, Ch. (1999). Analyse critique des indicateurs déterminant l'attribution des moyens destinés à la politique de discrimination positive en Communauté française de Belgique. Mesure et évaluation en éducation 22(2-3), 97-127. https://doi.org/10.7202/1091252ar

Dumay, X. et Dupriez, V. (2005). Effet établissement : quelles relations entre composition sociale des établissements et processus internes ? Mesure et évaluation en éducation, 28(2), 67-92. https://doi.org/10.7202/1087225ar

Duong Thi, D. T. et Loye, N. (2019). Analyses diagnostiques cognitives des résultats du test du Programme international de recherche en

lecture scolaire (PIRLS) 2011. Mesure et évaluation en éducation, 42(3), 29–69. https://doi.org/10.7202/1074103ar

Epskamp, S. (2023). Psychonetrics: Structural equation modeling and confirmatory network analysis. https://CRAN.R-project.org/package=psychonetrics

Epskamp, S., Maris, G., Waldorp, L. J. et Borsboom, D. (2018). Network psychometrics. Dans P. Irwing, T. Booth, et D. J. Hughes (dir.), The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development (p. 953–986). Wiley Blackwell. https://doi.org/10.1002/9781118489772.ch30

Falk, C. F., Huang, A. et Ilagan, M. J. (2025). Unsupervised [randomly responding] survey bot detection: In search of high classification accuracy. Psychological Methods. Advance online publication. https://doi.org/10.1037/met0000746

Fleck, L. (1979). Genesis and Development of a Scientific Fact. University of Chicago Press.

Fox, J.-P. (2010). Bayesian item response modeling: Theory and applications. Springer.

Gilbert, J. B. (2024). Modeling item-level heterogeneous treatment effects: A tutorial with the glmer function from the lme4 package in R. Behavior Research Methods, 56(5), 5055–5067. https://doi.org/10.3758/s13428-023-02245-8

Golino HF. et Epskamp S (2017) Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. PLoS ONE 12(6): e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Moulder, R., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Nesselroade, J., Sadana, R., Thiyagarajan, J. A. et Boker, S. M. (2021). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. Multivariate behavioral research, 56(6), 874–902. https://doi.org/10.1080/00273171.2020.1779642

Grove W. M. (2005). Clinical versus statistical prediction: the contribution of Paul E. Meehl. Journal of clinical psychology, 61(10), 1233–1243. https://doi.org/10.1002/jclp.20179

Hussein M.A., Hassan H. et Nassef M. (2019). Automated language essay scoring systems: A literature review. PeerJ Computer Science, 5, e208. doi: 10.7717/peerj-cs.208.

Jung, J. Y. (2025, 17 juillet) Improving automated scoring in reading assessments: Compress first, score next [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Kjell, O., Giorgi, S. et Schwartz, H. A. (2023). The text-package: An R-package for analyzing and visualizing human language using natural language processing and transformers. Psychological Methods, 28(6), 1478–1498. https://doi.org/10.1037/met0000542

Kuhn, T. S. (1962). The structure of scientific revolutions. University of Chicago Press.

Lafontaine, D. (1999). Un goût de lire bien mesuré. Élaboration d'une échelle d'attitudes envers la lecture. Mesure et évaluation en éducation, 22(1), 21-43.

Larose, R. et Racine, S. (1981). Des systèmes d'apprentissage opérationnels. Mesure et évaluation en éducation, 4(4), 83-105.

Laurencelle, L. (2015). Une théorie des seuils psychométriques à double contrôle d'erreur – Partie I : l'imprécision échantillonnale des centiles. Mesure et évaluation en éducation, 38(2), 87–110. https://doi.org/10.7202/1036764ar

Laurencelle, L. (2016). Une théorie des seuils psychométriques à double contrôle d'erreur – Partie II : l'erreur de mesure et le concept de norme sûre. Mesure et évaluation en éducation, 39(1), 45–66. https://doi.org/10.7202/1036705ar

Laurencelle, L. (2017). Une théorie des seuils psychométriques à double contrôle d'erreur – Partie III : les applications illustrées. Mesure et évaluation en éducation, 40(3), 105–133. https://doi.org/10.7202/1048912ar

Laveault, D. (2012). Soixante ans de bons et mauvais usages de l'alpha de Cronbach. Mesure et évaluation en éducation, 35(2), 1–7. https://doi.org/10.7202/1024716ar

Le-Bastard-Landrier, S. et Suchaut, B. (2000). L'élève comme objet d'évaluation au sein d'une démarche d'apprentissage de la lecture. Mesure et évaluation en éducation, 23(2-3), 21-36. https://doi.org/10.7202/1091225ar

Lin, T. Y., Tuerlinckx, F. et Vanbelle, S. (2025). Reliability for multilevel data: A correlation approach. Psychological methods, 10.1037/met0000738. Advance online publication. https://doi.org/10.1037/met0000738

Louis, R. (1982). La banque d'items : un outil pédagogique. Mesure et évaluation en éducation, 5(1), 22-30.

Louis, R. et Trahan, M. (1995). Une mesure des croyances des enseignants titulaires du primaire relative à trois approches d'évaluation des apprentissages. Mesure et évaluation en éducation, 17(3), 5-30.

Loye, N. (2005). Quelques nouveaux modèles de mesure. Mesure et évaluation en éducation, 28(3), 51-68. https://doi.org/10.7202/1087030ar

Loye, N. (2009). Les modèles cognitifs. Dans J.-G. Blais (dir.), Évaluation des apprentissages et technologies de l'information et de la communication : enjeux, applications et modèles de mesure. Presses de l'université Laval.

Loye, N. (2010). 2010, odyssée des modèles de classification diagnostique (MCD). Mesure et évaluation en éducation, 33(3), 75–98. https://doi.org/10.7202/1024892ar

Lu, Y. (2025, 16 juillet). Weighted likelihood estimator and its standard errors for sequential IRT models [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Madison, M. (2025, 16 juillet). Diagnostic classification models: Advancement through simplicity [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Merkle, E. (2025, 14 juillet). Bayesian psychometric modeling and blavaan [atelier]. IMPS 2025, Minneapolis, MIN, USA.

Mislevy, R. J. (2018). Sociocognitive foundations of educational measurement. Routeledge.

Ordre des psychologues du Québec. Introduction au dossier : lumière sur la psychométrie. https://www.ordrepsy.qc.ca/-/lumiere-sur-la-psychometrie

Racine, S. (1979). Une approche générale à la mesure de l'apprentissage. Mesure et évaluation en éducation, 2(2), 37-48.

Ramesh, D. et Sanampudi S.K. (2022). An automated essay scoring systems : A systematic literature review. Artif Intell Rev 55, 2495–2527. https://doi.org/10.1007/s10462-021-10068-2

Ren, H. (2025, 16 juillet). Use factor-augmented regularized latent regression to analyze complex large-scale assessment [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. Frontiers in Education, 5, Article 589965. https://doi.org/10.3389/feduc.2020.589965

Séguin, S. P. et Augier, R. (1986). Une introduction à la théorie des réponses aux items (TRI). Mesure et évaluation en éducation, 9(1), 7-44.

Sinharay, S. (2016). Asymptotically correct standardization of person-fit statistics beyond dichotomous items. Psychometrika 81, 992–1013. https://doi.org/10.1007/s11336-015-9465-x

Snijders T. (2001). Asymptotic distribution of person-fit statistics with estimated person parameter. Psychometrika, 66, 331-342.

Swaminathan, H. et Gifford, J. A. (1982). Bayesian estimation in the Rasch model. Journal of Educational Statistics, 7, 175-192.

Tutz, G. (1997). Sequential models for ordered responses. Dans W.J. van der Linden et R.K. Hambleton. (dir.) Handbook of Modern Item Response Theory. Springer. https://doi.org/10.1007/978-1-4757-2691-6_8

Valois, P. et Martin, R. (2008). Les modèles de mesure en éducation : enjeux, développements et orientations. Mesure et évaluation en éducation, 31(3), 125–153. https://doi.org/10.7202/1024995ar

van de Schoot, R. et Miočević, M. (2020). Small Sample Size Solutions (Open Access). Routledge. https://doi.org/10.4324/9780429273872

Zhili, S. (2025, 16 juillet). Model-based differential item functioning detection in cognitive diagnostic assessments [communication orale]. IMPS 2025, Minneapolis, MIN, USA.

Chapitre 11. Compétences non académiques : malgré les difficultés de conceptualisation et de mesure, un intérêt majeur pour les chercheurs et les praticiens en éducation

Sophie Morlaix

I. Introduction

La définition du concept de compétences non académiques n'est pas simple. Les termes pour les désigner sont nombreux (compétences sociales, socio-comportementales, psychosociales, socio-émotionnelles, relationnelles, savoir-être, soft skills…), et les disciplines s'y intéressant diverses (sciences de l'éducation et de la formation, économie, sociologie, psychologie, gestion…). L'apprentissage de ces compétences, comme leur nom l'indique, n'est pas forcément académique ou scolaire et ne relève pas d'une discipline spécifique, ni même d'une évaluation définie par des référentiels (Morlaix, 2023). Elles peuvent se rapporter à une dimension inter-individuelle (empathie, coopération, courtoisie…) et se développer dans la relation avec les autres, ou renvoyer à des dimensions plus intra-individuelles renvoyant à des comportements vis-à-vis de soi (persévérance, curiosité, estime de soi…).

Alors que les compétences dites académiques sont explicitement formulées dans les programmes scolaires, et leur évaluation envisagée par les enseignants, les compétences dites non académiques renvoient pour

leur part à des savoir-faire ou des comportements (savoir-être) plus implicites pour les élèves et leurs familles, voire pour l'école ou le système de formation. Ces compétences non académiques peuvent se développer dans différentes sphères : familiale, amicale, scolaire ou professionnelle. Elles pourraient s'apparenter à des « savoirs invisibles que l'école requiert et considère bien souvent comment allant de soi, sans guère travailler à en doter ceux qui n'en disposent pas, et que les enfants d'origine plus favorisée ont déjà en partie découverts et construits, avant et hors de l'école, grâce à l'environnement familial » (Rochex, 2008, p.162).

Les connaissances sur la définition du concept de compétences non académiques, sur leurs mesures et leurs évaluations manquent. En effet, les chercheurs comme les enseignants rencontrent un certain nombre de difficultés dans l'appréhension de ces différentes compétences : peu d'éléments empiriques sur ces compétences non académiques sont disponibles ; leurs mesures sont, de plus, très souvent des mesures déclaratives, sensibles à la formulation des questions et sans doute entachées de biais découlant de normes de désirabilité variables selon les contextes (par exemple, déclarer son goût pour la compétition peut être malvenu dans certains contextes culturels, Durut-Bellat, 2015).

L'objet de ce chapitre portera sur les éléments de conceptualisation des compétences non académiques à l'école et sur les difficultés de les mesurer notamment dans des contextes variés, mais pas uniquement. Il visera également à démontrer l'importance de leur prise en compte dans les travaux de recherche en éducation. En effet, même si leur mesure est souvent délicate, un certain nombre de travaux s'attachent aujourd'hui à appréhender ce concept tant son utilité pour expliquer les parcours scolaires ou professionnels et les inégalités qui les jalonnent, paraît importante.

II. Les compétences non académiques : des dénominations et des définitions multiples dans des recherches de plus en plus nombreuses et qui cherchent à les mesurer

L'apparition en 2015 de la notion de compétences non académiques dans le socle commun de connaissances et de compétences en France témoigne du rapport étroit que ces dernières semblent entretenir avec la réussite scolaire, comme le confirment certains travaux de recherche (Rayou, 1999 ;

Goodman et Gregg, 2010 ; Morlaix, 2015 ; Berthaud, 2019 ; Algan et Huillery, 2022 ; Lamboy et al, 2022…).

Les travaux qui s'intéressent à ce lien sont maintenant nombreux et émanent de disciplines et contextes différents. Par exemple, en 2001, les travaux de Heckman et Rubinstein montrent sur un échantillon de jeunes décrocheurs, qu'en dehors du champ académique, la motivation, la ténacité, le sérieux ou la persévérance sont des comportements tout aussi importants que de bons résultats scolaires pour s'insérer professionnellement. Il paraît étonnant, d'après ces auteurs, que les débats autour de la réussite se focalisent uniquement sur les capacités cognitives tout en laissant de côté les compétences dites non académiques. De la même façon, Borghans et al. (2006) abordent la question de l'impact des compétences non académiques sur les résultats aux tests cognitifs. Ces auteurs ont pour postulat que la performance réalisée lors d'un test cognitif ne dépend pas uniquement des capacités cognitives de l'individu mais également de certaines compétences dites non académiques comme l'application mentale, la volonté de procurer un effort mental supplémentaire afin de résoudre des problèmes, la persévérance… Des résultats similaires sont mis en évidence par Guimard et al. (2007) qui montrent dans leurs travaux l'importance de certaines compétences non académiques sur les performances scolaires. Ces dernières seraient influencées par la participation en classe, la coopération dans les situations d'apprentissage, l'intérêt pour les activités scolaires ou encore l'attention en classe.

Dans les travaux de Fanchini en 2016, l'identification d'une trentaine de compétences non académiques chez des élèves de primaire amène aux mêmes constats pour expliquer la réussite des élèves à l'école primaire. Les compétences non académiques sont déterminantes, ces dernières étant particulièrement corrélées aux caractéristiques personnelles des élèves (âge, sexe, catégorie socioprofessionnelle d'appartenance des parents). Ces constats, valables à l'école primaire le sont, quel que soit le niveau du système éducatif auquel on s'intéresse, même si dans l'enseignement supérieur, comme dans l'enseignement primaire ou secondaire, les travaux qui abordent directement les compétences non académiques en vue d'expliquer la réussite académique sont rares.

Au niveau du supérieur par exemple, une étude menée par Boyer et Sedlacek en 1987 montrait déjà que des variables non cognitives telles que la confiance en soi, l'autoévaluation réaliste, la recherche d'objectifs à long terme, la patience et la persévérance vis-à-vis des objectifs fixés, la tolérance ainsi que les connaissances et compétences développées dans des

domaines non académiques (par exemple service à la communauté ou expériences de leadership) sont des facteurs permettant d'obtenir de meilleures notes dans une population d'étudiants. De façon plus générale, la description du « métier d'étudiant » fait également état de diverses capacités ou comportements, qui ne sont pas tous de nature académique et qui s'apparentent à des compétences non académiques. Pour Alava (1999), en effet, l'acquisition du métier d'étudiant suppose d'acquérir des habiletés renvoyant à la fois au domaine cognitif mais aussi au domaine social, ces compétences ne devant pas être envisagées isolément les unes des autres mais plutôt dans leur dynamique d'intégration « personnelle et contextuelle » qui constitue une manière de faire, d'être et de savoir. Le métier d'étudiant requiert ainsi des capacités d'organisation, d'adaptation et de participation, ainsi que des capacités de gestion tant au niveau de l'activité studieuse à proprement parler qu'au niveau du temps libre, qui peut être alloué aux études ou à d'autres activités. Pour Beaupère et Grunfeld (2012) également, le métier d'étudiant requiert selon les publics une plus ou moins grande capacité d'adaptation, les étudiants n'ayant pas su s'adapter aux ruptures institutionnelles, sociales et académiques avec l'enseignement secondaire étant amenés à abandonner leurs études (Sarfati, 2013). Des entretiens menés par Boyer (2000) auprès d'étudiants entrant à l'université illustrent d'ailleurs ce propos : les étudiants qui décrivent une adaptation réussie mettent en avant leur capacité à « maîtriser les changements rencontrés », de même qu'ils associent cette adaptation réussie à « leur force de caractère et à leur indépendance » (p. 53). Ces résultats sont confirmés par ceux d'une recherche plus récente menée par Berthaud (2017) qui montre que certaines compétences non académiques liées à la socialisation étudiante, notamment l'initiative, le goût de l'effort et surtout le sentiment d'efficacité, apparaissent très significativement associés à de plus grandes chances de réussite et de meilleures notes finales, cet effet perdurant sous contrôle des autres déterminants.

Ces quelques exemples de travaux témoignent de la multiplicité et l'hétérogénéité des termes utilisés par les chercheurs intéressés par la notion de compétences non académiques. IIs ne permettent pourtant pas d'en dresser une liste exhaustive et cela pour diverses raisons. La première est qu'il existe une multitude de dénominations de ces compétences dites non académiques, selon les auteurs qui s'y intéressent, leur discipline de rattachement (économie, sociologie, psychologie, gestion, didactique, sciences de l'éducation...), ou encore leur statut (chercheurs, politiques, enseignants, membres du secteur associatif…), ou leur appartenance institutionnelle (universités, ministères, associations, entreprises…). Ainsi,

suivant les positionnements des acteurs, les dénominations varient : les organismes internationaux évoquent des compétences sociales et émotionnelles (OCDE, 2015), ou encore des compétences émergentes (Rocher, 2023), les travaux de recherche anglo-saxons invoquent le terme de soft skills (Heckman, 2012), des universitaires de l'école d'économie de Paris celui de compétences socio-comportementales (Huillery et Algan, 2022), le ministère de l'éducation nationale celui de compétences conatives (DEPP…), certains travaux en sciences de l'éducation ou en psychologie celui de compétences psychosociales (Tavant, 2021). Les travaux de l'IREDU (Institut de recherche sur l'éducation) qui s'intéressent depuis une quinzaine d'années maintenant à ce concept évoquent des compétences non académiques dans la mesure où elles ne figurent pas officiellement dans les programmes scolaires, ne bénéficient pas d'un temps scolaire spécifique et dédié pour les travailler, ne possèdent pas de référentiels d'évaluation officiels pour les mesurer par les enseignants et paraissent, de ce fait, différents des autres types de savoirs que l'École doit transmettre, qui sont explicités davantage dans les instructions officielles et pour lesquels des programmes d'objectifs et d'évaluation sont clairement définis. Ces différentes dénominations vont avoir pour conséquence de rendre difficile la définition même du concept. A nouveau, différentes raisons peuvent être évoquées pour comprendre cette difficulté. La première tient certainement à la multitude de référentiels nationaux et internationaux qui existent pour tenter de définir ces compétences non académiques qu'il serait impossible d'énumérer tant leur nombre est important (multiples référentiels métiers spécifiques à chaque branche, métiers, fonctions ; multiples référentiels de formation se rapportant à chaque diplôme, certification ; référentiels nationaux des compétences de base dans la sphère scolaire, extrascolaire, associative ; référentiels internationaux des compétences à acquérir pour être préparer au monde de demain…). Plusieurs tentatives infructueuses de l'OCDE en 1997 ou de l'IEA en 2019 ont d'ailleurs été envisagées pour établir des référentiels communs internationaux pour l'éducation, mais le renvoi de leur définition ou de leur importance à un contexte culturel et normatif rend difficile la définition et la délimitation d'une liste officielle de compétences non académiques à acquérir communément par les écoliers de tous les pays : par exemple, certains pays ou systèmes éducatifs comme ceux des pays émergents sont davantage préoccupés par le développement du goût de l'effort, ou de la compétition, alors que d'autres, plus au Nord de l'Europe, privilégient la coopération, l'entraide… En France, depuis 2022, une stratégie nationale visant à développer les compétences psychosociales des enfants et des jeunes (3-25 ans) a émergé : « Copilotée par les ministères en charge de la santé et de

l'Education nationale, cette stratégie interministérielle s'inscrit sur 15 ans (2022-2037) et associe neuf ministères (ministères en charge de la santé et de la prévention, de l'Education nationale et de la jeunesse, des lycées agricoles, de l'enseignement supérieur et de la recherche, de l'insertion professionnelle, de la justice, de la cohésion sociale, des sports, rejoints aussi par celui en charge de la culture). L'objectif est d'intégrer le développement des compétences psychosociales des enfants et des jeunes dans tous les milieux de vie et au sein de tous les apprentissages afin de permettre aux générations d'aujourd'hui et de demain de mieux réguler leurs émotions, vivre ensemble et prendre des décisions éclairées » (MEN, 2022).

Si l'intérêt des chercheurs pour l'analyse de ces compétences dites non académiques n'est plus à démontrer, de multiples interrogations se posent encore sur leur mesure et leur évaluation. Concernant leur mesure, la problématique repose principalement sur la difficulté d'opérationnaliser ces compétences non académiques. En effet, cette mesure repose souvent sur de l'autodéclaratif des apprenants ou des enseignants ou des représentations des uns ou des autres, non exempts de biais de désirabilité, ou se référant à des normes, ou de valeurs en vigueur dans les différentes sphères de socialisation (famille, école, système éducatif…). La mesure lors de mises en situation est particulièrement complexe notamment dans le cadre scolaire, le temps imparti à ces mises en situation n'étant pas extensible, la formation des enseignants certainement inadaptée, et les outils d'observations peu développés. Lorsqu'ils le sont, les outils peuvent être inadéquats par rapport au public auquel ils s'adressent, ou réservés à certaines disciplines (comme des tests psychologiques auxquels seuls les psychologues peuvent avoir accès), même si certaines échelles de mesure sont maintenant bien stabilisées (par exemple l'échelle de l'estime de soi développée par Harter en 1988, celle du sentiment d'efficacité personnelle développée par Bandura en 1997, l'état d'esprit de croissance de Dwek en 2010. Concernant la problématique de l'évaluation, plusieurs méthodes peuvent être citées (Solvet, 2023), comme les autoévaluations par les acteurs sur des échelles de mesure, ou l'observation en situation, ou des démarches types portfolios (pour accumuler des preuves ou des badges de la maîtrise de ces compétences). Pourtant, la difficulté qui caractérise ces méthodes est leur grande variété selon les contextes, le public auquel on s'adresse, ou encore le moment du parcours des individus… Ainsi, que l'on s'intéresse à la mesure ou à l'évaluation de ces compétences non académiques, un certain nombre de freins méthodologiques mériteraient d'être levés ou approfondis davantage si on souhaite que ce concept gagne en validité et fiabilité scientifiques.

III. Une complémentarité pourtant essentielle avec le développement des compétences cognitives

La relation positive entre compétences non académiques et réussite académique s'expliquerait en partie par la complémentarité qui peut exister entre les différents types de compétences nécessaires à la réussite du parcours de l'élève ou de l'étudiant. Ainsi, le rapport de l'OCDE de 2015 met en avant un point essentiel : le niveau actuel des élèves en compétences non académiques (sociales et émotionnelles) est plus important que le niveau de compétences cognitives pour le développement des futures compétences cognitives. Les enfants qui ont confiance en eux et qui croient en leurs capacités à jouer sur le futur (locus de contrôle interne), ou qui ont une capacité d'esprit de développement (growth mindset) ont plus tendance à atteindre des niveaux académiques élevés que ceux qui possèdent initialement uniquement de fortes compétences cognitives. Les recherches sur des données américaines vont dans le même sens : il existe un fort impact du niveau de quatre compétences cognitives, sociales et émotionnelles pendant l'enfance sur le développement des futures compétences (Cunha et Heckman, 2008, 2012).

De plus, les analyses montrent que le niveau accumulé de compétences non académiques (sociales et émotionnelles) joue un rôle important sur le développement des compétences cognitives, alors que le niveau de compétences cognitives, a un effet limité sur le développement des compétences sociales et émotionnelles. Un enfant avec un niveau élevé de compétences sociales et émotionnelles (par exemple : calme, respectueux et stable émotionnellement, dans le cas des résultats présentés) a plus tendance à réussir les tests cognitifs. Les résultats d'analyses issues de PISA 2012 sont d'ailleurs cohérents avec ces éléments : pour les élèves, leur rapport à l'école, leur croyance en leur réussite future et leur volonté de faire ce qu'il faut pour atteindre leurs objectifs, jouent un rôle central sur la maîtrise de sujets académiques et augmentent leurs capacités cognitives (OCDE, 2015). Ces constats démontrent la complémentarité entre les compétences cognitives et les compétences non académiques.

D'autres chercheurs en sciences de l'éducation ont travaillé sur cette notion de complémentarité, voire de substituabilité entre compétences académiques et compétences non académiques (Albandea et Giret, 2016 ; Giret et Morlaix, 2016). Ils montrent, dans certains cas, une substituabilité entre ces deux types de compétences, c'est-à-dire que les lacunes dans

certaines compétences académiques peuvent être compensées par des compétences non académiques, notamment dans les choix d'orientation scolaire ou pour expliquer les ambitions scolaires des élèves. En effet, de manière générale, les compétences non académiques affecteraient sensiblement les ambitions scolaires, directement et indirectement via le rapport à l'école. Une recherche menée en éducation prioritaire (Giret et Morlaix, 2016) met en avant le fait que les élèves se trouvant dans ces établissements ne sont pas dépourvus de ces compétences non académiques : le rapport à l'école et le sentiment de maîtrise sont des compétences non académiques particulièrement développées chez les élèves scolarisés en Réseau d'Education Prioritaire. Ainsi, ces résultats paraissent indiquer que les élèves scolarisés en éducation prioritaire se sentent plus en confiance par rapport à l'école et par rapport à leurs propres capacités que les élèves scolarisés hors éducation prioritaire, ce qui semble positif, notamment lorsque l'on examine les effets de ces variables sur l'ambition scolaire.

Pour Chauvel (2013) également, ces compétences non académiques occupent une place déterminante dans le processus d'orientation, notamment pour les établissements les moins favorisés, où l'attitude de l'élève vis-à-vis du travail, son autonomie et son comportement au sein de l'école et de la classe jouent un rôle important dans les décisions d'orientation venant des enseignants. Ces résultats font écho à ceux de Guillon et Huillery (2014) qui décomposent les facteurs qui structurent les inégalités sociales d'ambition scolaire, à niveau scolaire identique. Les aspirations scolaires sont le résultat de différents facteurs comme, par exemple, l'accès à l'information sur les possibilités d'orientation et les métiers, les contacts entre l'institution scolaire, la famille et les pairs mais plus largement d'un ensemble de représentations qui peut être affecté par des compétences non académiques, et dont les effets peuvent varier en fonction des milieux sociaux. La bienveillance des enseignants ou des équipes éducatives vis-à-vis des élèves contribuerait notamment à développer chez ces derniers, un certain nombre de compétences non académiques, favorables à leur réussite ou leur parcours scolaire. La bienveillance des enseignants renvoie à de nombreuses compétences non académiques en lien avec les apports conceptuels de Réto (2017) ou de Roelens (2019) : attention portée à autrui, écoute, disponibilité, compréhension, soutien émotionnel, mais également le non-jugement et l'acceptation d'autrui dont font état Shankland et al. (2018) dans leurs travaux.

IV. Au-delà des compétences non académiques développées par les élèves, l'importance de celles développées par les enseignants pour contribuer au processus d'apprentissage

Des travaux très récents, menés notamment à l'IREDU en 2021 et 2024 se sont intéressés à certaines compétences non académiques des enseignants et à leurs effets sur les élèves, notamment l'effet de la bienveillance des enseignants sur la réussite d'élèves de primaire ou d'étudiants à l'université a été, à plusieurs reprises, analysé à travers le développement de compétences non académiques chez ces derniers. Dans ces recherches, menées à des niveaux différents du système éducatif, sont mises en relation les compétences non académiques des enseignants et leurs effets sur le développement chez les apprenants de compétences non académiques. Ainsi les travaux de recherche de Morlaix et Tavant en 2021 visent à expliciter davantage l'effet enseignant mis en évidence dans les années 1990 au niveau des classes de primaire, par la mobilisation de ces nouvelles variables. L'objectif du travail, dans la lignée des travaux portant sur l'effet maître[1], est de mettre en évidence des profils interactionnels d'enseignants et de comprendre comment ce profil peut jouer sur le développement de compétences psycho-sociales ou non académiques chez les élèves (notamment six d'entre elles : l'empathie, la coopération, la maitrise de soi, l'estime de soi, l'internalité et l'anxiété). Les résultats proposés par cette recherche donnent des premiers éléments factuels intéressants, sur ce qui se joue entre l'enseignant et ses élèves au sein de la classe. Le comportement interactionnel de l'enseignant (perçu par les élèves à travers différentes dimensions qui peuvent s'apparenter à des compétences non académiques) semble lié au comportement psychosocial de l'élève, c'est-à-dire aux compétences non académiques développées par les élèves. La proximité perçue de l'enseignant ou à l'inverse la perception de son insatisfaction a des conséquences sur l'autoperception que l'élève a de ses compétences psychosociales. Ainsi, un élève qui ressentirait une forme de proximité chez son enseignant (empathique, responsabilisant, soucieux d'apporter de l'aide, mais également directif) pourrait

[1] L'effet maître est ici entendu en son sens le plus général : les progrès des élèves vont dépendre de manière significative de l'enseignant avec lequel ils sont scolarisés (Bresssoux, 1994).

développer plus de compétences non académiques en lien avec son bien être psychosocial. Ces premiers résultats ouvrent des pistes de réflexion intéressantes pour qui s'intéresse aux inégalités de réussite entre élèves, en mettant à jour le rôle de l'enseignant sur ces compétences psychosociales dont on sait maintenant qu'elles exercent une influence capitale pour expliquer les parcours scolaires et professionnels des individus.

Un autre exemple de recherche original peut être cité. Il s'intéresse au lien entre la bienveillance des enseignants dans le supérieur et les différentes formes d'engagement des étudiants en cours magistral (engagement cognitif, engagement affectif ou engagement comportemental). Ce travail de recherche (Morlaix et Duguet, 2024) a visé dans un premier temps à étudier les représentations des étudiants à l'égard de la bienveillance de leurs enseignants, et à mesurer dans un second temps les effets de ces représentations sur l'engagement dans les études. L'un des résultats les plus surprenants de cette recherche porte sur l'importance de la représentation de la bienveillance enseignante sur l'engagement affectif des étudiants. Ainsi, la représentation que les étudiants se font de la bienveillance de leur enseignant aurait l'effet le plus important sur la dimension de l'engagement qui pourrait paraître la moins importante dans l'enseignement supérieur : l'engagement affectif (et non sur l'engagement comportemental ou cognitif). Aussi paradoxal que puisse être ce résultat, puisque les étudiants déclarent quasiment pour un tiers d'entre eux qu'ils n'attendent pas forcément une bienveillance trop importante de la part de leurs enseignants, la dimension affective de cette bienveillance semble tout de même importante pour la majorité d'entre eux. Si cette dimension renvoie majoritairement au sentiment d'appartenance (Finn, 1989 ; Dost et Smith, 2023), ce résultat implique certainement une réflexion plus large sur l'importance de la socialisation étudiante à l'entrée dans l'enseignement supérieur, réflexion déjà engagée dans les travaux de Berthaud en 2017.

Ces résultats renvoient aux questions posées dans le rapport de synthèse du Grenelle de l'éducation en 2020 dans lequel les auteurs s'interrogeaient déjà sur les pratiques pédagogiques permettant de développer chez les élèves des compétences qu'ils appellent socio-comportementales. Les constats faits dans la recherche décrite ci-dessus invitent de la même façon à entamer une réflexion quant aux compétences professionnelles que les enseignants du supérieur se doivent aujourd'hui de développer pour favoriser la réussite étudiante.

Mais plus largement, ce sont les bénéfices à plus long terme du développement des compétences non académiques pour les élèves individuellement et pour la société dans son ensemble qui sont interrogés.

V. Compétences non académiques, marché du travail et bénéfices sociétaux

À l'instar des travaux relativement récents d'économistes de l'éducation (Bowles, Gintis et Osborne, 2001 ; Heckman et al., 2006) qui mettent en évidence l'importance de ces compétences non académiques pour l'individu sur le marché du travail, le rapport récent du conseil d'analyse économique (Guadalupe et Ng, 2022) attire l'attention sur l'importance du développement de ces *softs skills* chez les individus dès leur enfance pour l'économie française (la France étant plutôt en retard dans la dotation de ces compétences comparativement à d'autres économies développées). Ainsi, les bénéfices tant pour les individus à un niveau micro, ou ceux pour la société à un niveau plus macro, seraient particulièrement prégnants.

Au niveau micro, et individuel, les auteurs laissent entrevoir le développement d'une théorie comportementale des déterminants des salaires dans laquelle la productivité est déterminée non seulement par les compétences scolaires mais aussi par un ensemble de compétences non académiques, les employeurs accordant une place de plus en plus importante à ces dernières lors du recrutement d'une personne. Ces compétences constitueraient une sorte de capital culturel, témoignant d'un mérite professionnel (Duru-Bellat, 2009) complémentaire aux savoirs théoriques renvoyant au mérite scolaire.[2] En effet, avec la vive concurrence qui s'instaure sur le marché du travail, compte tenu du nombre croissant de diplômés de l'enseignement supérieur, les employeurs mobilisent d'autres critères de sélection et une gamme de compétences non académiques (Duru-Bellat, 2015). En s'appuyant sur une enquête du Céreq de 2008, Duru-Bellat rappelle que les « recruteurs utilisent effectivement le diplôme comme un signal de compétence académique minimal s'agissant de jeunes dépourvus d'expérience professionnelle. Mais ce signal est perçu comme insuffisant ; de fait, il se conjugue avec d'autres facteurs de nature

[2] La notion de capital culturel renvoie ici à des ressources culturelles non matérielles, mais intégrées à l'individu comme son éducation, son ouverture d'esprit…

variée, allant de la possibilité de bénéficier d'aides à l'emploi à des critères de motivation, à des façons de s'exprimer, de disponibilité, « voire de qualités personnelles très basiques (arriver à l'heure, être capable de travailler en équipe…) même chez les très diplômés » (Duru-Bellat, 2015).

Nealy montrait déjà en 2005 que les compétences techniques ne semblent plus être les seuls déterminants pour accéder à un emploi, en mettant en évidence le rôle de certaines compétences non académiques lors de l'embauche. Elles seraient d'ailleurs plus importantes que le nombre d'années d'études et joueraient un rôle particulier sur l'employabilité, concept qui renvoie à l'emploi et à la capacité d'être employé (Hillage et Pollard, 1998). De la même façon est maintenant mis en évidence un lien entre ces compétences non académiques et le niveau de rémunération des individus. Ainsi, certaines compétences transversales intra-individuelles (comportements vis-à-vis de soi) jouent positivement sur le niveau de salaire, comme en témoignent les travaux de Anderson et Bergman en 2001 qui mettent en lumière un lien entre persévérance et niveau de salaire, ou ceux de Goldsmith et al. (1997) mettant au jour un lien de cause à effet entre estime de soi et niveau de rémunération. Des liens statistiques sont maintenant établis entre compétences non académiques interindividuelles (comportements vis-à-vis des autres) et salaires. Ainsi, par exemple, le leadership jouerait positivement sur le salaire obtenu (Weinberger, 2014).

Une recherche menée par Albandea et Giret (2016) met à jour des effets différenciés de ces compétences non académiques sur les niveaux de rémunération, montrant ainsi que l'effet de la possession de compétences transversales ou non académiques est même plus important pour la distribution des hauts salaires, et pour accéder aux emplois les plus qualifiés, pour lesquels le niveau de diplôme n'est plus discriminant. Ainsi, et de façon plus générale, plusieurs travaux mettent en évidence les compétences non académiques recherchées par les employeurs. La méta-analyse réalisée par Robbes en 2012 sur le marché américain, est à cet égard très intéressante. Si on distingue les compétences non académiques intra-individuelles des compétences inter-individuelles, la première compétence non académique recherchée par les employeurs serait l'intégrité renvoyant à l'honnêteté, l'éthique, les valeurs de la personne et constituant une compétence non académique intra-individuelle. La deuxième serait une compétence de nature inter-individuelle, renvoyant aux compétences de communication, non seulement à la capacité de s'exprimer ou d'argumenter mais également à la capacité d'écouter. La troisième renvoie quant à elle à des normes de politesse, au savoir vivre et respect d'autrui. Sans poursuivre davantage la description des compétences non académiques recherchées en priorité par les employeurs,

cette classification a été mise à l'épreuve des faits dans une recherche menée en 2018 sur les diplômés de masters d'université française (Morlaix et Nohu, 2018). Les résultats montrent que plusieurs compétences non académiques expliquent la probabilité d'entrer sur le marché du travail ou d'être en emploi pour les étudiants de master : la capacité à travailler en équipe, la minutie, le dynamisme, l'innovation et la créativité paraissent être des compétences jouant positivement sur le fait d'être en emploi. Mais ces compétences, si elles paraissent utiles pour décrocher un emploi, ne sont pas forcément celles qui vont jouer sur les autres critères de l'employabilité que sont le statut de l'emploi, le niveau de l'emploi ou encore le salaire.

Une compétence non académique en particulier va jouer sur ces trois derniers critères et de façon positive : l'autonomie. Il s'agit d'une caractéristique que l'on peut supposer être développée dans les cursus de master notamment à l'université tant on connaît la grande liberté et autonomie pédagogiques laissées aux étudiants dans les cursus universitaires en général, dans ceux de master en particulier. Rappelons que cette compétence d'autonomie est l'une des deux les plus citées par les étudiants comme étant développées dans les cursus de formation en master. Elle constitue également une des compétences les plus utiles aux étudiants dans leur premier emploi. Le statut de l'emploi (CDD ou CDI, temps partiel…) est influencé, en plus de l'autonomie par le fait que l'individu soit minutieux, capable d'innover ou de résoudre des problèmes et témoigne d'une certaine éthique professionnelle, ces compétences non académiques jouant favorablement sur les aspects positifs du statut (plus forte probabilité d'avoir un CDI à temps complet…). Sur le niveau de l'emploi, et notamment la probabilité d'être embauché comme cadre pour les étudiants de master, la compétence de communication est la plus importante avec l'autonomie. Enfin, le niveau de rémunération va être influencé positivement non seulement par l'autonomie de l'individu mais également par son éthique professionnelle.

Au-delà des effets sur les trajectoires professionnelles individuelles, ce sont également les effets sociétaux du développement de ces compétences non académiques qui sont interrogés, même si les données sur le lien entre compétences non académiques d'une population et croissance économique ou effets sociétaux manquent, comme le met en exergue la note de synthèse de 2022 du conseil d'analyse économique. Des premiers résultats présentés au Conseil Scientifique d'Education Nationale (CSEN) par Algan et Huillery (2022) montrent, sur une cohorte québécoise (expérience dite de Montréal), les effets bénéfiques en termes d'externalités positives du fait d'avoir formé une cohorte de jeunes aux

compétences non académiques sur la vie future en société : baisse de 11% de la probabilité de commettre un acte criminel, baisse de 33% du nombre d'arrestations, et baisse de 40% du niveau de prestations sociales. Ces résultats mettent en lumière les externalités positives de l'éducation d'une population, connus en économie de l'éducation depuis les années 70. Ils sont cependant originaux, permettant de dépasser la vision traditionnelle du capital humain pour l'élargir à la possession de compétences non académiques.

VI. Pour conclure…

Le concept de compétences non académiques est difficile à définir, difficile à mesurer tant ses définitions peuvent être diverses et ses mesures variées. Pour autant, l'intérêt de sa prise en compte pour expliquer les parcours scolaires et professionnels ne fait plus de doutes. Les effets de ces compétences non académiques sur les parcours scolaires/universitaires et professionnels sont prouvés. Elles favorisent d'une part l'acquisition de compétences cognitives et académiques, qui définissent les situations de réussite et d'échec, et permettent de comprendre des inégalités de réussite entre élèves. De la même façon, sur le marché du travail, elles permettent aux individus de se démarquer et semblent liées notamment à une valorisation salariale, mais également à une adaptativité des parcours, ainsi qu'à des effets sociétaux importants. C'est pourquoi les avancées de la recherche sur les effets de ces compétences apparaissent primordiales pour saisir les processus à l'œuvre dans les différentes étapes et les choix de parcours scolaires et professionnels. Ainsi, des travaux à venir devraient investir davantage dans la définition et la mesure de ces compétences dites non académiques pour expliquer les inégalités de réussite, puis d'orientation aux différents niveaux du système éducatif.

VII. Quelques éléments de bibliographie

Algan Y. et Huillery E. (2022), *Economie du savoir-être*. Collection Sécuriser l'emploi. Presses de Sciences Po.

Albandea, I. et Giret, J.-F. (2016). L'effet des soft-skills sur la rémunération des diplômés. *Net.doc, CEREQ*, (149).

Duru-Bellat, M. (2015). Les compétences non académiques en question. Débats théoriques et interrogations empiriques. *Formation Emploi*, *130*(avril-juin), 13-29.

Duguet A. et Morlaix S. (2024), Les compétences psychosociales à l'université : entre bienveillance des enseignants et engagement des étudiants de licence en cours magistral, une analyse des liens. S*ymposium Recherche Education Formation (REF), 3-5 juillet 2024*, Université de Fribourg (Suisse).

Giret, J.-F. et Morlaix, S. (2016). *Les compétences sociales et non académiques dans les parcours scolaires et professionnels*. Editions Universitaires de Dijon.

Heckman, J. J., Stixrud, J. et Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market. *Journal of Labor Economics*, *24*(3), 411-482.

Morlaix, S. et Tavant D. (2021), Profil Enseignant et Compétences comportementales des élèves : Quels liens entre bienveillance de l'enseignant et bien être des élèves en classe ? *Revue Française de Pédagogie*, (211),87-101.

Morlaix, S. et Nohu, N. (2019), Compétences transversales et employabilité : de l'université au marché du travail. *Éducation permanente*, *218*(1), 109-118.

OCDE (2015). *Skills for social progress: the power of social and emotional skills*. OECD.

Weinberger, C. (2014), The Increasing Complementarity between Cognitive and Social Skills. *The Review of Economics and Statistics*, *96*(5), 849-861.

Chapitre 12. La menace de l'évaluation : 40 ans de recherche

Fabrizio Butera

I. Introduction

La boule au ventre. Les enfants qui partent le matin sachant qu'à l'école il y aura un test, les employées[1] qui quittent la maison pour aller au rendez-vous avec leur superviseuse lors du bilan annuel, les sportives qui se préparent pour une sélection, toutes décrivent ce moment d'anticipation d'une évaluation en évoquant la boule au ventre. Qui plus qui moins, certaines pas du tout. Pourquoi ? Être évalué est l'une des situations les plus courantes de la vie quotidienne - on peut ajouter la famille et les réseaux sociaux aux exemples ci-dessous qui portent sur l'école, le travail et le sport. Quelle est cette menace qui pèse sur nous dans la plupart des domaines de notre vie ?

Dans ce chapitre, nous aborderons de façon historique cette question, qui a occupé une bonne partie de notre activité de recherche ces 20 dernières années, ainsi que les travaux des 40 dernières années, qui ont fourni l'inspiration et la base théorique et empirique pour nos travaux.

[1] Dans ce texte, le féminin est utilisé comme genre générique.

II. Les précurseurs

Commençons par définir de quoi nous parlons quand nous nous référons à la menace de l'évaluation. Toute évaluation fournit une valeur à nos actions et productions. Nous pouvons évaluer l'intérêt d'un livre, la pertinence d'une dissertation, la vitesse d'une descente de slalom, la contribution d'un rapport financier, combien sont sages nos enfants, ou combien est élégante l'influenceuse que nous venons de voir sur l'un de nos réseaux sociaux. Ces évaluations peuvent être formelles (effectuées dans le cadre d'une fonction professionnelle d'enseignante, coach, manager ou autre) ou informelles (dans différents aspects de notre vie quotidienne), mais elles suivent des principes communs fondés sur un mécanisme important du fonctionnement humain : la comparaison sociale.

En 1954, le chercheur Léon Festinger a mis en évidence que nous sommes motivés à nous évaluer et que - à défaut d'instruments de mesure physique - nous le faisons en nous comparant aux autres. Par exemple, pour savoir si nous travaillons de façon assez efficiente, nous allons comparer notre travail à celui de nos collègues. Cette tendance a donné lieu à tout un ensemble d'instruments qui visent à formaliser ces comparaisons, dont le plus connu est le système de notation à l'école (Deutsch, 1975 ; 1979). Même si sa forme métrique (points, pourcentages…) fait penser à un système de mesure, l'attribution de valeur se fait par comparaison sociale : on a reçu une bonne note si la plupart des camarades ont reçu une note inférieure. On peut dire la même chose du chronométrage d'une course, pourtant fondé sur des instruments extrêmement précis : on gagne la médaille d'or non pas en fonction de la vitesse mesurée, mais à condition que les autres coureuses soient arrivées après.

Où rentre en ligne de compte la menace ? Cottrell, dans un chapitre de 1972, montre bien comment la simple présence d'autrui peut créer une appréhension d'évaluation. L'habitude d'être évaluée est telle, que quand on se sait observée par autrui, nous rentrons dans un état d'activation, anticipant une éventuelle évaluation de sa part. Mais ce n'est pas encore une menace. La menace intervient lorsque la comparaison sociale peut amener à constater notre infériorité. Bien sûr, on peut se comparer à des gens qui sont meilleurs ou plus avancés que nous pour en tirer des enseignements ou de l'inspiration (Festinger, 1954 ; voir Muller et Fayant, 2010, pour une discussion). Mais si cette infériorité a, ou risque d'avoir, des conséquences - l'autre gagne, ou est embauchée à notre place, ou devient

la chouchoute de la maîtresse, ou est plus visible - alors l'évaluation devient menaçante.

Pour résumer, avant les années 1980, plusieurs travaux ont mis en évidence que la comparaison sociale est un mécanisme fondamental du fonctionnement humain, et qu'il peut amener à une menace due à une infériorité constatée ou potentielle. Ceci est surtout vrai quand un système d'évaluation est en place et peut rendre visible cette infériorité, un système qui a des conséquences tangibles en termes d'accès à un diplôme, de statut au travail, ou d'appréciation dans un cercle social.

III. Les années 1980

Dans les années 1980, la littérature en psychologie sociale s'est attelée à définir la nature et l'origine de la menace de l'évaluation. Plusieurs théories ont mis en avant que nous avons besoin d'avoir une évaluation positive de nous-mêmes. Par exemple Deci et Ryan (1985) ont montré que le besoin de se sentir compétente est l'un des besoins psychologiques fondamentaux de l'être humain (voir aussi Ryan et Deci, 2000 pour une discussion). Au vu de cette importance, deux théories - la théorie de l'auto-affirmation de Steele (1988) et le modèle de la maintenance de l'auto-évaluation de Tesser (1988) - proposent que nous ayons une motivation fondamentale à réduire le décalage entre des standards qui sont importants pour nous (que nous avons fixés ou que d'autres ont fixés pour nous) d'une part, et l'évaluation que nous faisons de nous ou qui nous est renvoyée de l'extérieur d'autre part. Il apparaît donc que quand l'évaluation et les standards ne correspondent pas, ou risquent de ne pas correspondre, l'individu se sent profondément menacé (voir Tesser, 2001, pour une discussion). Par exemple, quand nous n'avons pas, ou risquons de ne pas avoir la moyenne que nos parents aimeraient que nous ayons à l'école, c'est là que la menace s'installe, rendue inéluctable par la visibilité des notes (Marshall et Weinstein, 1984).

Dans ce qui précède, nous avons insisté à plusieurs reprises sur le fait que la menace peut venir d'une évaluation qui n'est pas à la hauteur des standards ou attentes, mais aussi d'une évaluation qui n'est potentiellement pas à la hauteur des standards ou attentes. Dans leur travail sur l'interprétation des situations de stress, Lazarus et Folkman (1984) ont mis en évidence que cette interprétation peut avoir lieu en amont d'une situation dans laquelle on ne sait pas quelle sera l'issue, par exemple quel sera le résultat d'un examen. Selon ces auteurs, si l'individu

pense avoir plus de ressources que ce que demande la situation, alors une perception de défi s'ensuit. En revanche, si l'individu pense avoir moins de ressources que ce que demande la situation, alors une perception de menace s'ensuit.

En somme, dans les années 1980 les travaux sur la menace se sont attelés à définir sa nature et son fonctionnement. Pendant cette période, une attention particulière est portée aux mécanismes intra-individuels, la menace perçue. Cette tendance est cohérente avec le paradigme de la cognition sociale, paradigme dominant de cette époque en psychologie sociale (Fiske et Taylor, 1984).

IV. Les années 1990

En 1993, Pérez et Mugny publient le livre *Influences sociales : la théorie de l´élaboration du conflit*. Ce livre est important pour la question qui nous occupe dans ce chapitre, parce qu'il pose la question de la menace qui intervient dans les situations d'influence sociale. En effet, lorsque deux ou plusieurs personnes communiquent, discutent à propos de tâches ou problèmes communs, elles peuvent s'influencer mutuellement. C'est le cas, par exemple, de deux camarades qui discutent d'un problème de géométrie, ou d'une enseignante qui présente une leçon d'histoire à sa classe. La théorie de l'élaboration du conflit considère que l'issue d'une relation d'influence sociale dépend de comment les actrices de cette relation résolvent le désaccord qui peut émerger de la communication. En particulier, concernant les tâches de type scolaire ou académique - qui impliquent qu'une solution correcte existe - la résolution du conflit dépendra de la compétence perçue des deux (ou plusieurs) parties et de la menace qui est liée à la comparaison sociale entre les parties (voir aussi Mugny et al., 1996 ; 2001).

Les travaux menés par cette équipe montrent que les sources d'influence de haut statut ou de haute compétence sont perçues comme menaçantes, car leur compétence supérieure est vue comme impliquant l'infériorité de la cible d'influence (p. ex., Butera et al., 1993). Ceci a des conséquences importantes en termes d'influence sociale : craignant la supériorité de la source, la cible la suit, l'imite, acquiesce, mais sans vraiment traiter l'information qui est transmise dans la relation d'influence. Le manque de traitement en profondeur de l'information fait que la cible se limite à un changement de surface, avec peu de changements de fond ou à long terme (Butera et al., 1996 ; Maggi et al., 1996). On voit aisément que ces résultats

soulèvent un doute quant aux relations d'enseignement, où sont en présence une source d'influence de haut statut et de compétence supérieure (une enseignante) et des cibles de compétences moindres (des élèves). Les travaux menés dans les années suivantes permettront de préciser quelles sont les conditions qui permettent à une relation d'enseignement de produire des changements profonds et durables, comme nous le verrons plus loin.

Les sources d'influence de bas statut et de faible compétence, en revanche, ne sont pas vues comme menaçantes, car la comparaison avec la cible ne risque pas de montrer une infériorité de cette dernière. L'éventuel conflit peut alors focaliser l'attention sur la tâche ou sur le problème à traiter, et peut être élaboré de façon véritablement épistémique, en essayant de parvenir à une connaissance commune et nouvelle. Dans cette relation d'influence, l'imitation n'a pas vraiment de sens, mais les travaux montrent que l'élaboration épistémique du conflit peut amener à des changements en profondeur (Butera et Mugny, 1995 ; Maggi et al., 1996). La littérature sur le tutorat offre des résultats convergents, en montrant que, dans une relation entre partenaires qui n'ont que peu de différence de statut et de compétence, tutrice et tutorée s'engagent dans des échanges qui leur permettent d'aller en profondeur dans l'apprentissage (p. ex., Merrill et al., 1995).

Dans l'ensemble, ces travaux se focalisent sur les dynamiques interpersonnelles de la menace, en montrant comment les symétries et asymétries de compétence entre source et cible d'influence peuvent être plus ou moins menaçantes et avoir des conséquences au niveau de la profondeur du traitement de l'information.

V. Les années 2000

Les années 2000 amènent la confirmation que la comparaison avec un autrui supérieur est vraiment menaçante, à tel point qu'elle induit une distraction de la tâche en cours (Muller et al., 2004 ; Muller et Butera, 2007). Ces travaux démontrent aussi que la supériorité d'autrui n'a pas besoin d'être effective, mais que la simple possibilité qu'autrui soit supérieur est suffisante pour créer une menace. Dès les premiers degrés de l'école, nous apprenons que les différences d'évaluation ont des conséquences en termes de statut et de réussite, comme nous l'avons dit, ce qui transforme l'évaluation en un instrument au service de la compétition (Butera et al.,

2006). Cette habitude colore par la suite et de façon durable notre interprétation de l'évaluation et de la comparaison.

Un climat compétitif amène aussi les élèves et les étudiantes à se fixer des buts de performance, des objectifs fondés sur la démonstration de la compétence (Darnon et Butera, 2005 ; Elliot et McGregor, 2001). Si les élèves, sportives ou travailleuses évaluent de façon comparativement positive leur compétence, leurs buts de performance viseront à dépasser les autres - à être première de classe, médaille d'or ou employée du mois. Si le sentiment de compétence est comparativement faible, leurs buts de performance viseront à éviter l'échec ou une position excessivement défavorable. Plusieurs travaux de cette époque montrent que les buts de performance - comparativement aux buts de maîtrise, focalisés sur l'apprentissage et le progrès - promeuvent des modes relationnels de régulation du dialogue et du désaccord avec autrui (Darnon et al., 2006 ; 2007). Autrement dit, des étudiantes, par exemple, guidées par des buts de performance auront tendance à essayer de déterminer qui a raison et qui a tort dans un débat sur un sujet scolaire, plutôt que de creuser le sujet pour en étudier et découvrir ses multiples facettes (Darnon et Butera, 2007).

La focalisation sur la performance amène les apprenantes à considérer que la compétence est exclusive, ou en tout cas que les mesures de performance, comme les notes à l'école, permettent de déterminer qui est la plus compétente des élèves. Comment dépasser cette interprétation conflictuelle des compétences ? Dans la décennie précédente, il avait déjà été montré que c'est l'interdépendance négative (ce que l'une gagne, l'autre le perd, la compétition) qui empêche les gens en interaction de se mettre d'accord (Butera et Mugny, 1995). Dans les années 2000, on applique un raisonnement complémentaire : pour se mettre d'accord, il faut travailler dans un contexte d'interdépendance positive (le groupe atteint ses buts si tous les membres atteignent les leurs, la coopération). En suivant les enseignements de la tradition de recherche sur l'apprentissage coopératif (p. ex., Johnson et Johnson, 2009), un programme de recherche a montré que lorsqu'on travaille dans une situation d'interdépendance positive des buts, mais aussi des ressources (ce qui rend les membres d'un groupe complémentaires), les relations à l'intérieur du groupe s'améliorent (Buchs et Butera, 2001) et la compétence d'autrui, loin d'être une menace, devient un prédicteur d'apprentissage (Buchs et al., 2004 ; Buchs et Butera, 2009).

Dans cette période, les travaux se sont donc attachés à mettre en évidence comment le contexte - plus compétitif ou plus coopératif - peut amener à considérer autrui comme étant une menace pour le soi.

VI. Les années 2010

Le terme « menace de l'évaluation » commence à être utilisé de façon programmatique avec la sortie du livre *L'évaluation, une menace ?* (Butera et al., 2011). Ce livre réunit plusieurs collègues travaillant en Suisse, France, Belgique et Canada et propose un panorama diversifié et international de la recherche sur les conséquences potentiellement menaçantes de l'évaluation. Il est à la fois l'aboutissement et le point de départ d'une communauté qui étudie les effets de l'évaluation de façon critique et multidisciplinaire, réunissant la psychologie sociale, la psychologie cognitive, les sciences de l'éducation, la sociologie et les études genre.

La première partie est définitionnelle et fait état des composantes de la menace de l'évaluation. La comparaison sociale est reconnue comme le mécanisme à la base de la menace et les dangers de la sous-évaluation de la compétence de soi sont mis en évidence (voir aussi Butera et Darnon, 2017). Concernant les origines structurelles, l'utilisation systématique des notes est analysée comme une pratique qui traverse tous les domaines de l'activité humaine, et qui crée une habitude à la menace de l'évaluation (voir aussi Hayek et al., 2015). Un contrat didactique clair et fondé sur la transmission de connaissances est identifié comme étant un élément qui peut atténuer la menace de l'identité.

Dans la deuxième partie, les auteurs et autrices ont abordé le contexte de la transmission des connaissances, l'un des contextes les plus importants dans les systèmes éducatifs. Ici, le rôle de la comparaison sociale est analysé en tenant compte des structures scolaires et des pratiques éducatives. Les dynamiques de classe, le rôle des attentes des enseignantes et les pratiques scolaires sont analysés comme sources possibles de menace de l'identité des élèves. L'importance de la compétition qui teinte la comparaison sociale avec autrui est largement reconnue (voir aussi Butera et al., 2018). Même les pratiques pédagogiques alternatives, comme l'apprentissage coopératif, sont décortiquées pour mettre en évidence sous quelles conditions elles peuvent contrecarrer ou accentuer la menace (voir aussi Buchs et al., 2011).

La troisième partie analyse comment la fonction de sélection des systèmes éducatifs peut accentuer la vulnérabilité des élèves à la menace (voir aussi Sommet et al., 2013), et en particulier des élèves qui sont déjà en situation de vulnérabilité à cause de leur classe sociale d'origine (voir aussi Smeding

et al., 2013), leur genre (voir aussi Souchal et al., 2014) ou leur origine ethnique ou culturelle.

Finalement, la quatrième partie met en évidence que la menace de l'évaluation peut avoir des conséquences qui vont bien au-delà de l'apprentissage et de la réussite scolaire. La menace peut amener à un ensemble de comportements antisociaux, de l'agression à la rétention d'information (voir aussi Toma et al., 2013 ; Toma et Butera, 2015) et à la tricherie (voir aussi Pulfrey et Butera, 2013).

Ce livre, et la recherche qu'il a organisée et générée (dont nous avons indiqué quelques références), ouvre donc la voie à l'étude des conséquences de la menace de l'évaluation. Cette étude englobe un large éventail de niveaux d'analyse, de la motivation aux buts, de l'apprentissage aux comportements antisociaux.

VII. Les années 2020 et le futur

Après des années passées à étudier la menace de l'évaluation pour les élèves et les étudiantes, les années 2020 ouvrent une nouvelle phase d'attention aux effets de la menace de l'évaluation représentée par certaines pratiques des enseignantes (p. ex. Butera et al., 2020). À vrai dire, une série d'études avait déjà commencé à sortir en 2019. D'une part, plusieurs études montrent que l'utilisation de l'évaluation normative amène les enseignantes à perpétuer les inégalités de classe sociale existantes, en évaluant plus favorablement les élèves de milieu social favorisé que les élèves des classes populaires (Autin et al., 2019). D'autre part, des études montrent que la sélection scolaire amène aussi les enseignantes à perpétuer les inégalités de classe sociale existantes, en orientant vers des filières d'études longues les élèves de milieu social favorisé plus que les élèves des classes populaires (Batruch et al., 2019).

Un autre courant de recherche s'est penché sur le climat de classe induit par les enseignantes, avec une attention particulière au climat de l'erreur, qui peut être plus positif - en considérant que l'erreur est la base de l'apprentissage - ou plus négatif - interprété comme une faute (Steuer et al., 2013). En particulier, une étude montre que les enseignantes peuvent avoir un contrôle sur le climat de l'erreur induit dans leur classe (Soncini et al., 2021). Ceci est particulièrement important, car le climat de l'erreur que perçoivent les élèves peut avoir un impact sur leurs résultats scolaires, avec un climat plus négatif qui réduit l'apprentissage (Soncini et al., 2022). La raison pour laquelle ces résultats sont pertinents pour la question de la

menace de l'évaluation est qu'il a été montré que, bien que l'erreur puisse être considérée comme une base fondamentale pour l'apprentissage, lorsque les enseignantes pensent à l'erreur dans le cadre de l'évaluation, alors elle est interprétée comme une faute (Soncini et al., 2023).

La tendance de ces dernières années est donc d'intégrer de plus en plus la complexité des variables structurelles dans la caractérisation de la menace de l'évaluation, notamment la fonction sélective de l'école et le climat compétitif qui s'ensuit (Darnon et al., 2009 ; Butera et al., 2021). Cette tentative d'intégration est justifiée par notre travail récent sur le modèle de la circulation de la compétition dans les systèmes éducatifs (Butera et al., 2024a). Dans ce travail, nous avons montré comment les idéologies, valeurs et normes liées à la compétition transitent de la société vers les établissements d'éducation, quel que soit leur niveau. Ici, nous avons mis l'accent sur des idéologies et valeurs comme la méritocratie, la foi dans un marché libre équitable et le néolibéralisme, ainsi que sur des normes telles que l'accent mis sur la productivité et l'employabilité. Ensuite, nous avons passé en revue les dispositifs et environnements compétitifs présents dans les institutions éducatives qui influencent les aspirations, les attitudes et les comportements des élèves, notamment à travers des dispositifs tels que l'évaluation normative, la sélection scolaire ou encore le *numerus clausus*, ainsi que des variables contextuelles comme le climat de classe, le cadre de fixation des objectifs et les attitudes face à l'erreur. Finalement, nous avons mis en lumière que cette cascade d'influences sociales peut avoir un effet sur les valeurs, objectifs et comportements compétitifs des élèves, ce qui peut se répercuter sur leurs résultats scolaires, mais aussi sur la qualité des interactions sociales en classe. Nous avons fermé la boucle en discutant les conséquences de cette forme de socialisation des élèves sur la société, qu'il s'agisse de la perpétuation des structures existantes ou d'une possible transformation sociale. Dans les travaux futurs, nous espérons pouvoir étudier les formes que prend la menace des compétences dans le passage à travers un tel cercle d'influences sociales (voir aussi Butera et al, 2024b). Il serait aussi important d'étudier le rôle des décisions politiques, au niveau national et dans les politiques éducatives, sur le maintien d'une culture de la menace des compétences.

VIII. Bibliographie

Autin, F., Batruch, A. et Butera, F. (2019). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology, 111*, 717–735.

Batruch, A., Autin, F., Bataillard, F. et Butera, F. (2019). School selection and the social class divide: How tracking contributes to the reproduction of inequalities. *Personality and Social Psychology Bulletin, 45*, 477–490.

Buchs, C. et Butera, F. (2001). Complementarity of information and quality of relationship in cooperative learning. *Social Psychology of Education, 4*, 335- 357.

Buchs, C. et Butera, F. (2009). Is a partner's competence threatening during dyadic cooperative work? It depends on resource interdependence. *European Journal of Psychology of Education, 24,* 145-154.

Buchs, C., Butera, F. et Mugny, G. (2004). Resource in(ter)dependence, student interactions and performance in cooperative learning. *Educational Psychology, 24,* 291-314.

Buchs, C., Gilles, I., Dutrévis, M. et Butera, F. (2011). Pressure to cooperate: Is positive reward interdependence really needed in cooperative learning? *British Journal of Educational Psychology, 81,* 135–146

Butera, F., Batruch, A., Pulfrey, C., Autin, F. et Toma, C. (2020). Les effets de la menace dans l'évaluation normative. Dans N. Younès, C. Gremion et E. Sylvestre (dir.), *L'évaluation, source de synergies ?* (p. 53-68). Editions de l'ADMEE-Europe.

Butera, F., Batruch, A., Autin, F., Mugny, G., Quiamzade, A. et Pulfrey, C. (2021). Teaching as social influence: Empowering teachers to become agents of social change. *Social Issues and Policy Review, 15,* 323-355.

Butera, F., Buchs, C. et Darnon, C. (dir.) (2011). *L'évaluation, une menace ?* Presses Universitaires de France.

Butera, F. et Darnon, C. (2017). Competence assessment, social comparison and conflict regulation. Dans A. Elliot, C. Dweck et D. Yaeger (dir.), *Handbook of Competence and Motivation (2e éd., Theory and Application*, p. 192-213*)*. Guilford Press.

Butera, F., Darnon, C., Buchs, C. et Muller, D. (2006). Les méfaits de la compétition : comparaison sociale et focalisation dans l'apprentissage. Dans R.V. Joule et P. Huguet, (dir.), *Bilans et perspectives en psychologie sociale* (p. 15-44). Presses Universitaires de Grenoble.

Butera, F., Dompnier, B. et Darnon, C. (2024b). Achievement goals: A social influence cycle. *Annual Review of Psychology*, *75*, 527-554.

Butera, F., Świątkowski, W. et Dompnier, B. (2024a). Competition in education. Dans S. Garcia, A. Tor et A. Elliot (dir.), *The Oxford handbook on the psychology of competition* (p. 569–597). Oxford University Press.

Butera, F., Legrenzi, P. et Mugny, G. (1993). De l´imitation à la validation : études sur le raisonnement. Dans J. A. Pérez et G. Mugny (dir.), *Influences sociales : la théorie de l´élaboration du conflit* (p. 99-120). Delachaux et Niestlé.

Butera, F. et Mugny, G. (1995). Conflict between incompetences and influence of a low-competence source in hypothesis testing. *European Journal of Social Psychology, 25,* 457-462.

Butera, F., Mugny, G., Legrenzi, P. et Pérez, J.A. (1996). Majority and minority influence, task representation, and inductive reasoning. *British Journal of Social Psychology, 35,* 123-136.

Butera, F., Sommet, N. et Toma, C. (2018). Confirmation as coping with competition. *European Review of Social Psychology*, *29*, 299-339.

Cottrell, N., B. (1972). Social facilitation. Dans C. G. McClintock (dir.), *Experimental social psychology* (p. 185-236). Holt.

Darnon, C. et Butera, F. (2005). Buts d'accomplissement, stratégies d'étude, et motivation intrinsèque : présentation d'un domaine de recherche et validation française de l'échelle d'Elliot et McGregor (2001). *L'Année Psychologique, 105,* 105-131.

Darnon, C. et Butera, F. (2007). Learning or succeeding? Conflict regulation with mastery or performance goals. *Swiss Journal of Psychology, 66,* 145-152.

Darnon, C., Butera, F. et Harackiewicz, J. (2007). Achievement goals in social interactions: Learning within a mastery vs performance goal. *Motivation and Emotion, 31,* 61-70.

Darnon, C., Dompnier, B., Delmas, F., Pulfrey, C. et Butera F. (2009). Achievement Goal Promotion at University: Social Desirability and Social Utility of Mastery and Performance Goals. *Journal of Personality and Social Psychology, 96,* 119-134.

Darnon, C., Muller, D., Schrager, S., Pannuzzo, N. et Butera, F. (2006). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology, 98*, 766-776.

Deci, E. L. et Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* Plenum.

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues, 31*, 137–149.

Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, *34*, 301-401.

Elliot, A. J. et McGregor, H. A. (2001). A 2*2 Achievement Goal Framework. *Journal of Personality and Social Psychology, 80*, 501-519.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Fiske, S. T. et Taylor, S. E. (1984). *Social cognition*. Addison-Wesley.

Hayek, A.S., Toma, C., Oberlé, D. et Butera, F. (2015). Grading hampers cooperative information sharing in group problem solving. *Social Psychology, 46,* 121–131.

Johnson, D. W. et Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*, 365-379.

Maggi, J., Butera, F. et Mugny, G. (1996). Conflict of incompetences: direct and indirect influences on representation of the centimetre. *International Review of Social Psychology, 9,* 91-105.

Marshall, H. H. et Weinstein, R. S. (1984). Classroom factors affecting students self-evaluations: An interactional model. *Review of Educational Research, 54,* 301-325.

Merrill, D. C., Reiser, B. J., Merrill, S. K. et Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and instruction*, *13*(3), 315-372.

Muller, D., Atzeni, T. et Butera, F. (2004). Coaction and upward social comparison reduce the illusory conjunction effect: Some support for distraction-conflict theory. *Journal of Experimental Social Psychology, 40,* 659-665.

Muller, D. et Butera, F. (2007). The distracting effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology, 93,* 194-211.

Muller, D. et Fayant, M. P. (2010). On being exposed to superior others: Consequences of self-threatening upward social comparisons. *Social and Personality Psychology Compass*, *4*, 621-634.

Mugny, G., Butera, F. et Falomir, J.M. (2001). Social influence and threat in social comparison between self and source's competence: Relational factors affecting the transmission of knowledge. Dans F. Butera et G. Mugny (dir.), *Social influence in social reality* (p. 225-237). Hogrefe & Huber.

Mugny, G., Butera, F., Sanchez-Mazas, M. et Pérez, J. A., (1995). Judgements in conflict: The conflict elaboration theory of social influence. Dans B. Boothe, R. Hirsig, A. Helminger, B. Meier et R. Volkart (dir). *Perception – Evaluation - Interpretation. Swiss Monographs in Psychology* (Vol. 3, p.160-168). Hans Huber.

Pérez, J.A. et Mugny, G. (dir.) (1993). *Influences sociales : la théorie de l´élaboration du conflit* (p. 99-120). Delachaux et Niestlé.

Pulfrey, C. et Butera, F. (2013). Why neo-liberal values of self-enhancement lead to cheating in higher education: A motivational account. *Psychological Science, 24,* 2153–2162.

Ryan, R. M. et Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68-78.

Smeding, A., Darnon, C., Souchal, C., Toczeck-Capelle, M.C. et Butera, F. (2013). Reducing the socio-economic status achievement gap at university by promoting mastery-oriented assessment. *PLoS ONE, 8*(8): e71678. doi:10.1371/journal.pone.0071678

Sommet, N., Pulfrey, C. et Butera, F. (2013). Did my M.D. really go to university to learn? Detrimental effects of *numerus clausus* on self-efficacy, mastery goals and learning. *PLoS ONE 8*(12): e84178. doi:10.1371/journal.pone.0084178

Soncini, A., Matteucci, M. C. et Butera, F. (2021). Error handling in the classroom: an experimental study of teachers' strategies to foster positive error climate. *European Journal of Psychology of Education, 36*, 719-738.

Soncini, A., Matteucci, M.C. et Butera, F. (2023). Errors: Springboard for learning or tool for evaluation? Ambivalence in teachers' error-related beliefs and practices. *Social Psychology of Education*.

Soncini, A., Visintin, E. P., Matteucci, M. C., Tomasetto, C. et Butera, F. (2022). Positive error climate promotes learning outcomes through students' adaptive reactions towards errors. *Learning and Instruction*, *80*, 101627.

Souchal, C., Toczek-Capelle, M.C., Darnon, C., Smeding, A., Butera, F. et Martinot, D. (2014). Assessing does not mean threatening: Assessment as a key determinant of girls' and boys' performance in a science class. *British Journal of Educational Psychology, 84,* 125–136.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. Dans L. Berkowitz (dir.), *Advances in experimental social psychology* (Vol. 21, p. 261–302). Academic Press.

Steuer, G., Rosentritt-Brunn, G. et Dresel, M. (2013). Dealing with errors in mathematics classrooms: Structure and relevance of perceived error climate. *Contemporary Educational Psychology*, *38*, 196-210.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. Dans L. Berkowitz (dir.), *Advances in experimental social psychology.* (Vol. 21, p. 181–227). Academic Press.

Tesser, A. (2001). *On the plasticity of self-defense. Current Directions in Psychological Science, 10,* 66–69.

Toma, C. et Butera, F. (2015). Cooperation versus competition effects on information sharing and use in group decision making. *Social and Personality Psychology Compass, 9*, 455–467.

Toma, C., Vasiljevic, D., Oberlé, D. et Butera, F. (2013). Assigned experts with competitive goals withhold information in group decision making. *British Journal of Social Psychology, 52,* 161–172.

Chapitre 13. Comment « évaluation » et « didactique » s'articulent-elles dans deux champs de recherche anglophones majeurs : les Pedagogical Content Knowlegde et l'Assessment Literacy ?

Évaluation et didactique dans le champ des PCK et de l'AL

Céline Lepareur, Annick Fagnant et Raphaël Pasquini

I. Introduction

Dans un récent rapport du CNESCO issu de la conférence de consensus sur les pratiques d'évaluation en classe, Fagnant (2023) défend, à la suite de Benett (2011) et de Brookhart (2011), l'idée selon laquelle les compétences professionnelles en évaluation des enseignants et enseignantes pourraient être améliorées au travers de programmes de formation qui mobilisent à la fois les apports des travaux récents en matière d'Assessment for Learning (AfL - Allal et Laveault, 2009) et les connaissances pédagogiques de contenu (Pedagogical Content Knowledge - PCK) qui sont impliquées dans l'enseignement d'une discipline spécifique.

En examinant les multiples catégories de connaissances professionnelles des enseignant·es, le concept de PCK introduit par Shulman (1986) a

suscité un grand intérêt en invitant à porter une attention spécifique aux contenus enseignés. En revanche, excepté dans de rares travaux, l'attention portée aux pratiques disciplinaires a été largement négligée dans la recherche en évaluation (Cowie et Moreland, 2015). Pour pallier ce manque, notre contribution propose d'articuler deux cadres théoriques anglophones majeurs : celui des PCK qui catégorise les connaissances pédagogiques de contenu dont l'enseignant·e a besoin pour enseigner une discipline donnée, et celui de l'Assessment Literacy (AL[1]), défini comme un ensemble de connaissances et de compétences dont l'enseignant·e a besoin pour mettre en œuvre efficacement l'évaluation en classe (Pastore, 2023). Nous faisons en effet l'hypothèse que ces deux cadres gagneraient à être pensés conjointement afin de mieux comprendre la façon dont les connaissances des enseignant·es s'opérationnalisent dans les pratiques d'évaluation en classe, en fonction des disciplines et, *in fine*, impactent les apprentissages des élèves.

Bien que l'intérêt du concept de PCK soit largement reconnu pour catégoriser les connaissances des enseignant·es, la manière de les modéliser prête encore aujourd'hui à débat (Depaepe et al., 2013, 2025 ; Park et Suh, 2019 ; Şen, 2023). À ce titre, le cadre de l'Assessment Literacy apparaît pertinent, non seulement pour progresser dans la compréhension de ces modélisations, mais aussi pour dépasser une opposition entre didactique et évaluation à la lumière des PCK. Il épouse en effet une épistémologie située (Pasquini et Morales Villabona, 2022), en cohérence avec l'AfL, en donnant une importance majeure aux disciplines dans les pratiques d'évaluation, et il considère que le développement des pratiques rend compte d'un processus dynamique, toujours en construction, notamment pour des questions de spécificités disciplinaires.

La contribution interrogera les articulations entre évaluation et didactique en donnant un aperçu de quelques travaux ayant problématisé cet objet. Le champ des recherches mobilisant le cadre des PCK étant très vaste, nous faisons le choix de focaliser notre texte sur deux disciplines scolaires : les sciences et les mathématiques. Ce choix résulte du fait que les modèles développés dans ces deux disciplines sont à la fois les plus foisonnants et qu'ils permettent d'illustrer aisément la place variable occupée par l'évaluation dans ces derniers. Puis, en se focalisant sur l'évaluation en classe, le cadre théorique des PCK sera mis en synergie avec celui de l'AL.

[1] Ce terme a été traduit en français par « littératie en évaluation » (Pasquini et Morales Villabona, 2022), mais au vu de ses origines anglophones, nous resterons fidèles à son appellation première.

Dans une perspective de développement professionnel, ces deux cadres considèrent que c'est l'interaction complexe entre diverses connaissances évaluatives et des connaissances propres à l'enseignement d'un objet disciplinaire spécifique (PCK) qui participe au développement de l'identité d'évaluateur·rice d'un·e enseignant·e (Abell et Siegel, 2011 ; Xu et Brown, 2016). Ainsi, et sans prétention de retracer l'historique des recherches dans ce domaine, la contribution s'efforcera de présenter quelques modèles fondateurs, tout en cherchant à mettre en lumière des pistes de recherches futures dans ce domaine encore en construction.

II. La place variable de l'évaluation dans les modèles PCK pour l'enseignement des sciences et des mathématiques

L'objectif poursuivi à travers l'élaboration du concept de PCK était, pour Shulman (1986), de faire reconnaître le métier d'enseignant·e comme une profession à part entière, fondée sur des savoirs spécifiques construits en situation. L'auteur soutient en effet qu'enseigner ne se limite pas à transmettre un savoir disciplinaire, mais qu'il convient aussi de le rendre accessible ; l'enseignant·e doit être en mesure de le légitimer, de le situer dans un cadre épistémologique et de le rendre intelligible pour les élèves. Situées à l'interface entre les connaissances disciplinaires et les connaissances pédagogiques générales, les PCK semblent impacter la qualité de l'enseignement et l'apprentissage des élèves (voir Depaepe et al., 2013, 2020 pour des synthèses en mathématiques et van Driel et al., 1998, 2014 ; Kind, 2009 pour les sciences).

Le modèle de Shulman a depuis servi de socle au développement d'un champ de recherche à l'échelle internationale. Dans sa publication de 1987, sept types de connaissances sont mis en évidence : 1) connaissance des contenus disciplinaires 2) connaissance pédagogique générale 3) connaissance pédagogique de contenu 4) connaissance du curriculum 5) connaissance des apprenant·es et de leurs caractéristiques 6) connaissance des contextes éducatifs 7) connaissance des buts de l'éducation, des valeurs éducatives et leurs fondements historiques et philosophiques. La manière dont ces connaissances interagissent entre elles et la façon dont les enseignant·es les mobilisent en contexte d'enseignement ont, par la suite, fait l'objet de foisonnantes recherches. Si cette catégorisation a favorisé une approche originale des connaissances enseignantes, ni strictement disciplinaire ni purement pédagogique, sa simplicité a

toutefois conduit plusieurs chercheur·es à interroger les relations entre les catégories de connaissances (Cross, 2025 ; Hill et al., 2008). Bien que le modèle initial de Shulman s'inscrive dans une vision généraliste, la majorité des travaux qui s'en sont emparés se concentrent dans les disciplines des sciences et des mathématiques (Depaepe et al., 2013 ; Loughran et al. 2012). Nous présentons alors successivement quelques modèles majeurs dans ces deux disciplines en vue de dégager leurs spécificités, tout comme leurs points de convergence.

II.a. Des composantes de PCK contrastées dans les modèles pour l'enseignement des sciences

Grossman (1990) fut l'une des premières à s'emparer de la catégorisation des connaissances de Shulman (1987) pour la réorganiser en quatre catégories : les connaissances sur le contenu disciplinaire à enseigner, les connaissances pédagogiques générales, les connaissances du contexte et les PCK. Chacune de ces catégories se décline ensuite en un ensemble de composantes. La composante « *Assessment* » n'est alors pas encore formellement intégrée aux catégories de connaissances des enseignant·es. Il faudra attendre les propositions développées de manière quasi-simultanées par Veal et Makinster (1999) et Magnusson et al. (1999), pour que l'évaluation occupe une place à part entière dans les modèles de PCK.

Focalisons-nous d'abord sur le modèle de Magnusson et al. (1999) qui fut très largement mobilisé dans les recherches ultérieures. Celui-ci (Figure 1) définit les PCK au travers de cinq composantes, chacune étant constituée de sous-composantes. Parmi elles, on y distingue la composante « *Knowledge of Assessment of Scientific Literacy* » qui décrit les connaissances en évaluation de la littératie scientifique, dont disposent les enseignant·es de sciences. Cette composante se subdivise en deux sous-catégories distinguant les dimensions de l'apprentissage scientifique à évaluer (ce

qu'il faut évaluer) des connaissances des méthodes pour les évaluer (comment les évaluer).

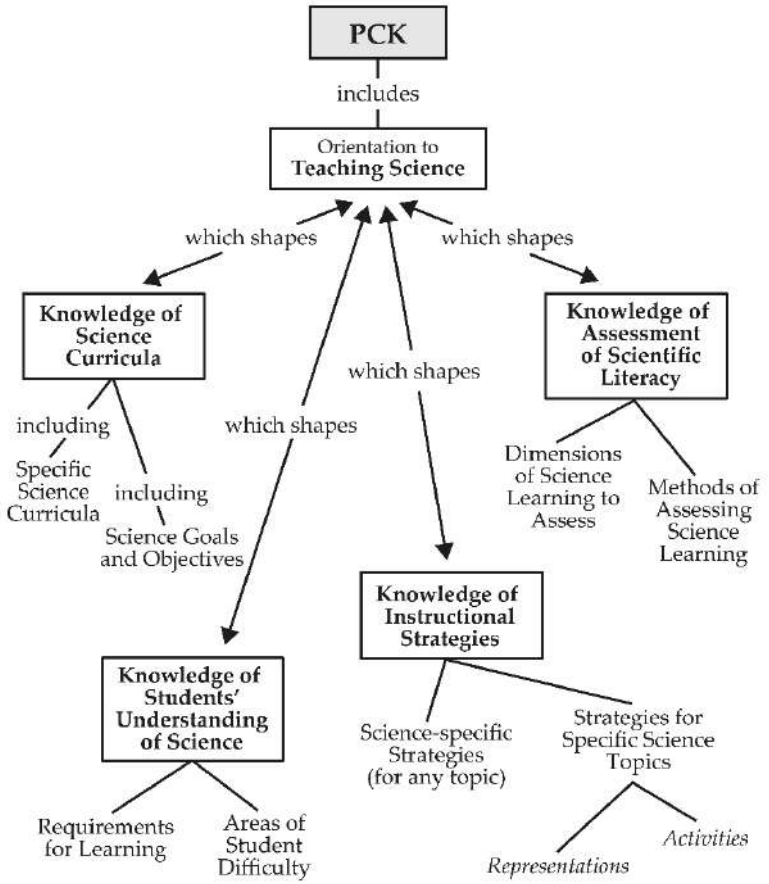


Figure 1 : Composantes de PCK pour l'enseignement des sciences, extrait de Magnusson et al. (1999), p. 99

Bien que cette catégorisation des PCK ait été largement reprise dans les travaux menés dans le champ, la linéarité des relations entre les composantes, tout comme la place de « *Orientation to teaching science* » qui chapeaute le tout, a été fortement remise en question (Friedrichsen et al., 2011). Park et Oliver (2008) ont ainsi proposé une nouvelle modélisation dans laquelle les cinq composantes des PCK s'influencent toutes mutuellement. Ils ont également introduit l'idée fondatrice selon laquelle les PCK des enseignant·es se développent en interaction avec les composantes au travers de deux processus : l'intégration (lorsque l'enseignant·e pense à son enseignement) et la réflexion-en-action (en référence aux ajustements constants de l'agir en classe). Il s'agissait ainsi de montrer la relation dynamique entre les différentes connaissances détenues par l'enseignant·e et celles qui se développent dans l'action.

Toutefois, notons que la composante « *assessement* » n'a pas encore été véritablement investiguée par les chercheur·es en didactique des sciences pour bien comprendre de quelle évaluation il s'agit, et ce qu'elle implique en termes de gestes professionnels. Certain·es préciseront par la suite que comprendre les connaissances des enseignant·es en matière d'évaluation ne se limite pas à repérer les stratégies ou les contenus mobilisés dans l'évaluation, mais que cela implique également de s'intéresser à la manière dont ils et elles interprètent les résultats des évaluations et agissent en fonction (Abell et Siegel, 2011 ; Edwards, 2020). Il est intéressant de noter que cette dernière préoccupation est aussi au cœur de l'Assessment Literacy, sur laquelle nous revenons plus loin.

Le concept de PCK a fait l'objet de nombreuses conceptualisations, et la divergence des points de vue entre les chercheur·es rend délicate la comparaison des résultats de recherche (Abell, 2008). C'est notamment pour surmonter cette difficulté qu'une conférence de consensus, visant à réunir les spécialistes de la communauté des PCK en sciences, s'est tenue en 2012 afin d'avancer sur une définition commune et ainsi clarifier les méthodes permettant de les appréhender. Le modèle développé à l'issue de cette conférence est présenté en Figure 2.

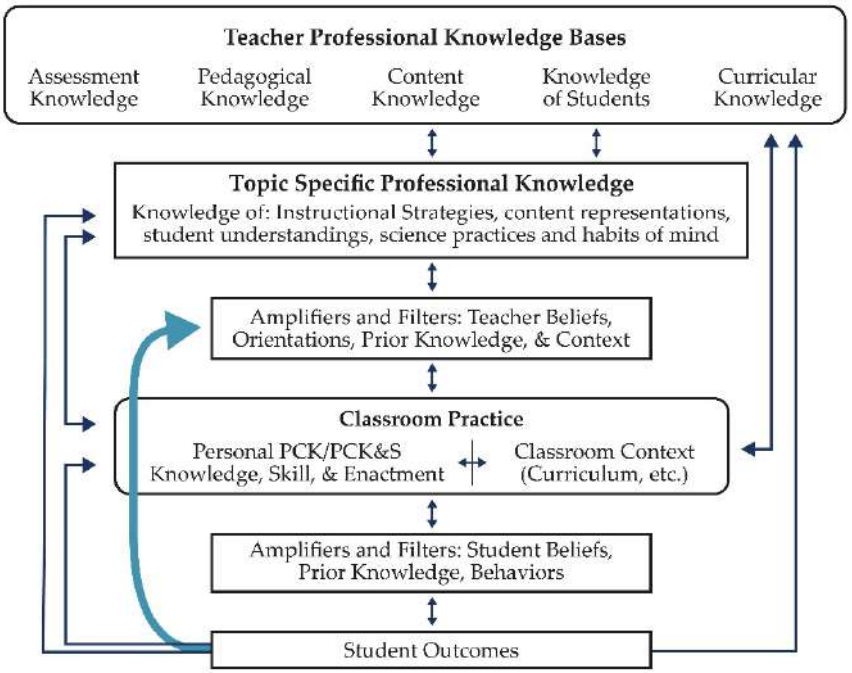


Figure 2 : Modèle issu de la 1ère conférence de consensus de 2012 ; extrait de Gess-Newsome et al., 2015, p. 29

Ce modèle marque un nouveau virage quant à la place accordée à l'évaluation. Celle-ci devient en effet une des cinq connaissances de base, considérées dans ce modèle comme des savoirs de référence, produits par la recherche, dans lesquels les enseignant·es viendraient puiser. Ce modèle met ainsi en exergue la personnalisation des savoirs en connaissances qui passent à travers le « prisme » des enseignant·es ainsi que le rôle des amplificateurs et des filtres qui traduisent les connaissances professionnelles des enseignant·es dans leurs pratiques effectives. À ce titre, ce modèle s'inscrit dans une perspective dynamique, que l'on retrouvera dans l'AL : il prend en compte les points de vue, croyances, expériences, réactions affectives des enseignant·e·s ainsi que leur contexte, qui agissent comme autant d'amplificateurs lorsqu'ils les amènent à accorder plus d'attention à leurs actions, ou comme des filtres, lorsqu'ils les poussent au contraire à les réduire, voire à les rejeter.

Bien que soulignant une dimension dynamique des connaissances en jeu, ce modèle a cependant essuyé nombre de critiques : des divergences d'interprétation des « *Topic-specific professional knowledge* » ont notamment émergées au sein même de la communauté de chercheur·es. Une seconde conférence de consensus, tenue en 2017, a proposé à sa suite le « *refined consensus model* » (Carlson et al., 2019) pour simplifier les catégories et délimiter les différents types de PCK selon qu'elles sont partagées par une communauté (*collective PCK*), personnelles aux enseignant·es (*personnal PCK*), ou mises en œuvre (*enacted PCK*). Cette modélisation, accessible en annexe 1, rend plus explicite la nature des savoirs et connaissances, et replace l'enseignant·e à la fois comme individu mais aussi comme élément d'un groupe.

En résumé, alors qu'une bonne part des modèles développés en didactique des sciences ont placé l'évaluation en tant que composante des PCK, les deux modèles de consensus placent l'évaluation dans la catégorie des connaissances de base détenues par l'enseignant·e. Qu'en est-il dans le champ disciplinaire des mathématiques ?

II.b. La place discrète de l'évaluation dans les modèles de PCK en mathématiques

Il y a une dizaine d'années, Depaepe et al. (2013) ont proposé une revue systématique des travaux menés sur les PCK dans le champ des mathématiques. Cette équipe de recherche a mis en évidence que la modélisation la plus influente était celle proposée par Hill et al. (2008 ; voir aussi Ball et al., 2008) : le modèle MKT (*Mathematical Knowledge for Teaching*). Dans une revue narrative ultérieure, Depaepe et al. (2020) se sont intéressé·es aux liens entre PCK et qualité de l'enseignement des mathématiques et ont notamment pointé un autre modèle majeur, le modèle COACTIV (*COgnitively ACTIVating instruction)* développé par Kunter et al. (2013).

Le modèle MKT s'appuie sur les travaux de Shulman (1987) et développe plus spécifiquement trois des sept types de connaissances que l'auteur avait alors mis en évidence : les connaissances de contenu, les connaissances pédagogiques de contenu et la connaissance du curriculum (cf. Figure 3). Un des apports de ce modèle est d'avoir distingué plusieurs catégories aux « connaissances de contenu » et aux « connaissances pédagogiques de contenu » (PCK).

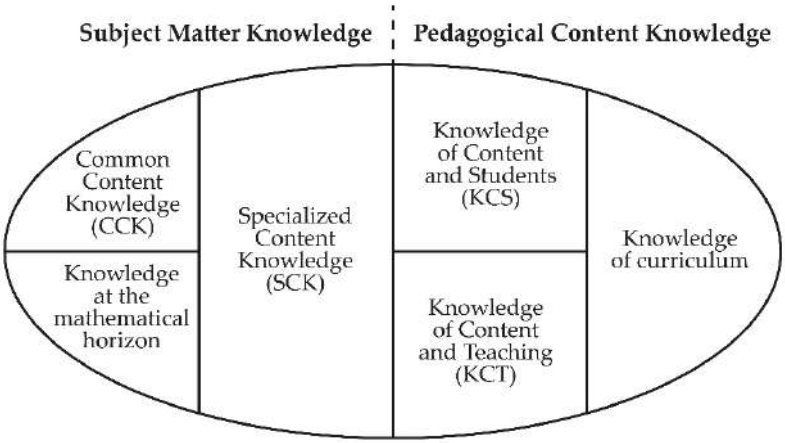


Figure 3 : Le modèle Mathematical Knowledge for Teaching (MKT) - Extrait de Hill et al. (2008), p. 377

Dans ce modèle, les « connaissances de contenu » comprennent trois sous-catégories : 1) Être soi-même capable d'effectuer les exercices que l'on va travailler avec les élèves (*Common content knowledge*) ; 2) Connaître des aspects mathématiques particuliers au métier d'enseignant·e (par ex. être capable de résoudre un même problème de plusieurs façons - *Specialized content knowledge*) ; 3) Savoir situer les thématiques mathématiques de l'année où l'on enseigne par rapport à celles qui ont été vues avant et celles qui seront vues après (*Horizon content knowledge*). Les PCK recouvrent quant à elles : 1) les connaissances propres aux élèves (par ex. comprendre les erreurs des élèves ainsi que les concepts qui leur posent particulièrement problème - *knowledge of content and students*) ; 2) les connaissances centrées sur l'enseignement (par ex. disposer d'un répertoire d'exemples parlants et sélectionner des activités intéressantes pour introduire un nouveau contenu - *knowledge of content and teaching*) et 3) les connaissances des programmes scolaires en vigueur (*knowledge of content and curriculum*).

On peut noter d'emblée que le terme « *assessment* » ne se retrouve explicitement dans aucune des dénominations des sous-catégories identifiées par les auteur·rices. Il paraîtrait toutefois erroné de considérer que l'évaluation est totalement absente puisqu'elle trouve en réalité sa place tant dans les « *knowledge of content and teaching* » (par ex. sélectionner un exercice approprié pour évaluer la compréhension des nombres

décimaux) que dans les « *knowledge of content and students* » (par ex. interpréter les erreurs commises par des élèves qui réalisent des soustractions écrites)[2].

Dans le modèle COACTIV, présenté en Figure 4, les trois sous-catégories de PCK recouvrent partiellement le modèle précédent (on y retrouve les versants *student* et *teaching*) tout en y ajoutant une sous-catégorie (les *explanatory knowlege*). Baumert et Kunter (2013) décrivent ces dernières comme faisant notamment référence au potentiel didactique des tâches, à leur exigence cognitive et à la planification de l'enseignement de contenus spécifiques. Les *kowledge of mathematical tasks* sont notamment celles qui vont permettre de fournir des explications et des représentations multiples pour faire comprendre les notions enseignées aux élèves. Enfin, les *knowledge of students mathematical thinking* portent sur les conceptions erronées, erreurs typiques et stratégies variées développées par les élèves, ainsi que sur les « *ways of assessing student knowledge and comprehension processes* » (p. 33), faisant ainsi explicitement mention à l'évaluation.

[2] Ces exemples proviennent d'items libérés par Ball et al., 2008 ; Hill et al., 2008 et disponibles sur https://lmt.marsal.umich.edu/files/lmt_sample_items.pdf

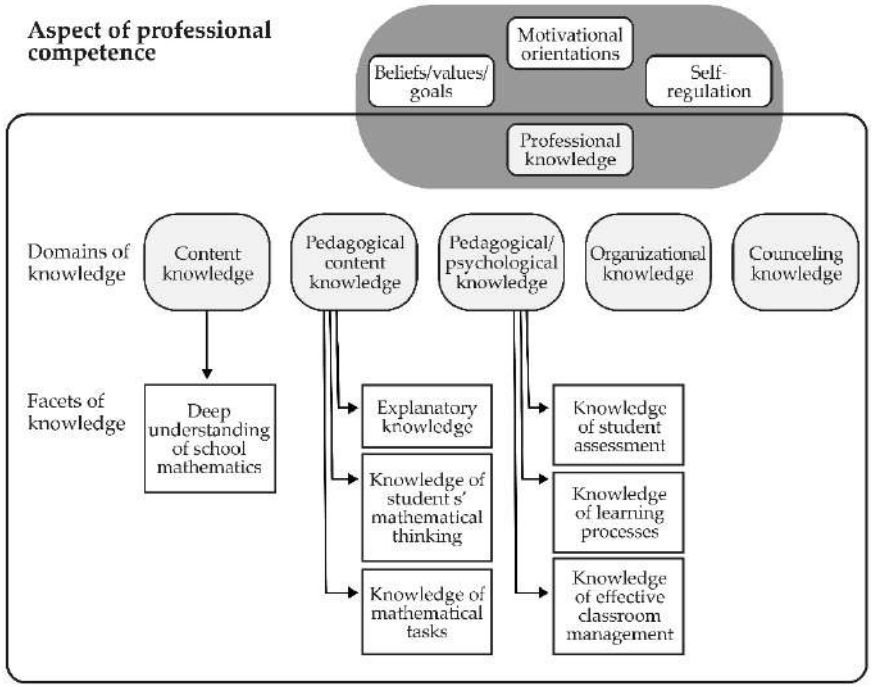


Figure 4 : Le modèle COACTIV des connaissances professionnelles - Extrait de Baumert et Kunter, 2013, p. 29

Le modèle COACTIV est bien plus large que le modèle MKT : il cherche à décrire la compétence professionnelle et inclut notamment différents types de connaissances, dont les connaissances psychopédagogiques générales. L'évaluation y est explicitement mentionnée comme une sous-dimension principale constituée elle-même de sous-catégories que nous ne détaillerons pas ici (voir Baumert et Kunter, 2013) mais qui ne sont pas sans rappeler celles que l'on retrouve dans les travaux sur l'Assessment Litteracy développés au point III.

Tout comme en sciences, la façon de décomposer les connaissances professionnelles varie selon les modèles. Ainsi, le modèle à la base de l'étude *TEDS-M* menée par l'IEA[3] propose trois catégories de

[3] TEDS-M (pour Teacher Education and Development Study in Mathematics) Voir https://www.iea.nl/studies/iea/teds-m

connaissances professionnelles : les connaissances de contenu, les connaissances pédagogiques générales et les PCK (Kaiser et Bloëmeke, 2013). Comme dans le modèle COACTIV, l'évaluation est une sous-catégorie des connaissances pédagogiques générales et n'est pas identifiée comme une sous-catégorie des PCK. Toutefois, une analyse plus détaillée (Blömeke et Delaney, 2012) de ce que recouvrent ces sous-catégories fait apparaître des éléments propres à l'évaluation, aussi bien dans les *curricular knowledge* (eg., *knowing about different assessment formats*) que dans les phases préactive (eg., *choosing assessment formats*) et interactive (eg., *analyzing or evaluating students' mathematical solutions or arguments*) qui constituent les trois sous-dimensions des PCK de ce modèle (Kaiser et Bloëmeke, 2013).

Le dernier exemple retenu est la modélisation proposée par Chapman (2015) dans le domaine spécifique de la résolution de problèmes. S'appuyant sur le modèle MKT, l'auteure distingue classiquement les connaissances de contenu et les PCK (avec les deux sous-dimensions « *student* » et « *teaching* »), auxquelles s'ajoutent des facteurs affectifs et des croyances. Dans des travaux ultérieurs, son équipe propose un questionnaire opérationnalisant les deux versants des PCK et met à jour plusieurs sous-dimensions parmi lesquelles une spécifiquement dédiée à l'évaluation en résolution de problèmes (Piñeiro et al., 2021).

Ce tour d'horizon ciblé de quelques modèles propres aux mathématiques illustre que l'évaluation relève d'une connaissance (psycho)pédagogique générale et qu'elle n'est généralement pas identifiée au premier niveau de catégorisation des PCK, comme s'il ne s'agissait pas d'un champ majeur de celles-ci. L'évaluation apparaît toutefois de façon plus ou moins explicite lorsque l'on investigue davantage ce que recouvre chacune de ces sous-catégories. À la lumière de ces constats, l'Assessment Literacy permet de prolonger la réflexion.

III. L'inscription des PCK dans l'Assessment Literacy

L'Assessment Literacy (AL) fait l'objet de nombreux travaux dans la communauté scientifique anglophone depuis le milieu des années 1990. Cette perspective théorique est née du besoin de mieux comprendre quelles connaissances et compétences sont nécessaires à un·e enseignant·e pour qu'il·elle puisse différencier une évaluation « fondée » d'une évaluation « non fondée » (Stiggins, 1995). Son ancrage dans une

perspective de développement professionnel est donc premier. L'intérêt que suscite l'AL provient de trois phénomènes relativement récents : les pressions qui pèsent de plus en plus sur les pratiques évaluatives des enseignant·es et les responsabilités qui en découlent ; la nécessité de soutenir l'apprentissage des élèves en développant des pratiques d'évaluation adaptées au sein de leurs classes ; l'impact des disciplines sur les pratiques d'évaluation rendant toute généralisation risquée.

Pastore (2023), dans une revue de littérature portant sur 42 études récentes, a montré que les cadres conceptuels de l'AL ont également beaucoup évolué depuis ses débuts. Plutôt que de retracer cette évolution, il nous semble plus pertinent ici de mettre en exergue deux éléments : tout d'abord, le fait que les PCK – même si elles n'ont pas été explicitement mentionnées dans les diverses modélisations réalisées - ont depuis longtemps été considérées comme une composante phare de l'AL (voir à ce titre les travaux de Brookhart, 2011) ; et d'autre part, l'idée que l'AL accorde aujourd'hui une attention aux aspects sociaux, contextuels et théoriques en jeu dans les pratiques d'évaluation, à l'instar de certaines modélisations des PCK précédemment évoquées.

Ainsi, l'AL est consensuellement définie comme « une pratique sociale dynamique dépendant du contexte qui implique que les enseignant·es articulent et négocient les connaissances scolaires et culturelles entre elles et eux, et avec les élèves, dans l'initiation, le développement et la pratique de l'évaluation pour atteindre les objectifs d'apprentissage » (Willis et al., 2013, p. 242, trad. libre). L'épistémologie située de l'AL souligne que les pratiques évaluatives s'appréhendent dans des contextes spécifiques de niveaux micro (classes, élèves), méso (écoles, équipes pédagogiques) et macro (directives légales, relations aux parents) qui interagissent entre eux (Fulmer et al., 2015). Conjointement, l'AL postule que les démarches d'évaluation s'inscrivent dans les processus d'enseignement et d'apprentissage, eux aussi toujours contextualisés, notamment du point de vue des disciplines (Pasquini et Morales Villabona, 2022). Aussi, les différentes composantes de l'AL ne se juxtaposent pas, mais représentent bien davantage un réseau de connaissances en interactions et contextualisées, participant dans une perspective systémique au développement des compétences en évaluation de l'enseignant·e.

Cela étant, plusieurs modélisations de ces connaissances en interactions co-existent aujourd'hui dans la recherche contemporaine. Certaines, comme celles d'Abell et Siegel (2011), Pastore et Andrade (2019) ou Pastore (2023), soulignent les multiples négociations que les enseignant·es sont amené·es à réaliser entre différents facteurs dans leurs prises de décisions

en évaluation. Ces modélisations sont avant tout systémiques et ne hiérarchisent pas les connaissances entre elles. Mais une question les réunit : à quelles conditions mieux comprendre comment les enseignant·es intègrent de nouvelles visées de l'évaluation dans leur pratique, comment ils.elles transfèrent leurs apprentissages professionnels en classe, notamment du point de vue de la pédagogie mise en œuvre ? L'épistémologie située et contextuelle de l'AL montre qu'il existe différentes dispositions chez l'enseignant·e dans sa capacité à développer sa littératie en évaluation (Pastore, 2023). La modélisation de l'AL telle que conceptualisée par Xu et Brown (2016) est à ce titre éclairante pour aborder notre problématique, dans la mesure où elle accorde une place prépondérante aux PCK.

III.a. Une modélisation de l'AL où les PCK jouent un rôle central

En articulant les apports de 100 recherches issues du champ de l'évaluation et de la formation des enseignants, Xu et Brown (2016) proposent une conceptualisation de l'AL à des fins de développement professionnel (Figure 5).

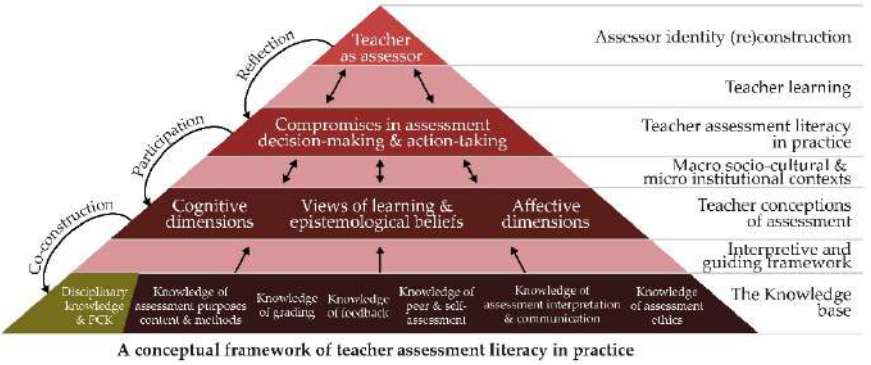


Figure 5 : La modélisation TALiP (Xu et Brown, 2016, p.19)

Nommée *Teacher Assessment Literacy in Practice* (TALiP), cette conceptualisation met en synergie plusieurs composantes. D'abord, les auteurs considèrent qu'une base de connaissances, au nombre de sept, sont essentielles : des connaissances didactiques et pédagogiques de contenu

(PCK) ; des connaissances sur les buts, contenus et méthodes d'évaluation ; des connaissances en notation ; en feedback ; sur l'auto-évaluation et l'évaluation par les pairs ; sur l'interprétation et la communication de l'évaluation ; sur l'éthique en évaluation. Ces connaissances entrent en interaction avec les conceptions que l'enseignant·e a de la cognition, de l'apprentissage, de l'épistémologie de sa discipline et de l'impact des émotions en évaluation, comme autant d'éléments qui jouent le rôle de filtres interprétatifs. Notons ici qu'une interaction apparaît particulièrement pertinente pour notre propos : celle qui concerne les PCK de l'enseignant·e avec sa conception de sa discipline et de son épistémologie. Un processus de co-construction de connaissances entre en jeu, ces dernières pouvant faire l'objet d'interprétations variées.

L'AL de l'enseignant·e prend alors forme dans une perspective dynamique, dans la mesure où ces connaissances, réinterprétées par l'enseignant·e, passent au filtre de ses différents contextes (équipe, école, système éducatif), comme dit précédemment. Cette littératie en développement se concrétise notamment dans une participation active à des actions de formations diverses, où les prises de décisions, actions et compromis que l'enseignant·e réalise en évaluation sont débattus. Ainsi, l'AL est éminemment située, puisque les disciplines scolaires et les contextes infléchissent son développement. Enfin, Xu et Brown (2016) montrent à quel point ce que l'enseignant·e apprend en formation entre en jeu dans le développement de son AL, jusqu'à influencer son identité d'évaluateur·rice, comprise comme le résultat d'une réflexion continue sur les différentes dimensions de son AL en construction. À ce titre, les conditions de cet apprentissage se concrétisent dans une perspective idéalement collaborative de développement professionnel.

Cette conceptualisation se distingue d'autres précédemment évoquées. En effet, bien qu'interactive, elle est également étapiste, dans la mesure où l'AL ne peut ici se développer sans l'appropriation de connaissances de base. Nous pouvons toutefois y lire un parallèle entre les connaissances « mises en œuvre » (cf. 2è modèle de consensus, Carlson et al., 2019) et le niveau de la prise de décision tel que les auteurs le formalisent. Ensuite, elle permet de penser les conditions et modalités de formation à l'évaluation dans lesquelles les PCK ont une place prépondérante. Ainsi, et même si l'AL d'un individu n'est jamais stabilisée une fois pour toutes, l'appropriation progressive et continue des PCK par les enseignant·es et leur mise en interaction avec les autres composantes est fondamentale.

III.b. Un exemple de recherche croisant PCK et AL

Sur la base de leurs travaux empiriques en didactique des sciences d'une part et du croisement des modèles de Magnusson et al. (1999) et de Pellegrino et al. (2001) d'autre part, Abell et Siegel (2011) proposent un modèle de l'Assessment Literacy (cf. Figure 6). Pour ces auteures, ce modèle se destine non seulement à la formation des enseignant·es, mais également à la recherche dans ce domaine encore peu exploré. Elles proposent dès lors un classement des connaissances en évaluation comme suit : la connaissance des finalités de l'évaluation, qui renvoie aux raisons pour lesquelles les enseignant·es choisissent d'évaluer (pourquoi évaluer ?) ; la connaissance de ce qu'il faut évaluer, qui concerne les contenus scientifiques et les domaines que les enseignant·es choisissent d'évaluer (quoi évaluer ?) ; la connaissance des stratégies d'évaluation, qui désigne les méthodes utilisées pour évaluer les apprentissages des élèves (comment évaluer ?) ; la connaissance de l'interprétation et de la prise de décision, qui se rapporte à la manière dont les enseignant·es analysent les données issues des évaluations et les actions qu'ils et elles en tirent (quelles décisions ?). Comme on peut le voir, les auteures placent au centre du modèle le rôle déterminant des conceptions de l'apprentissage, mais aussi des valeurs et principes dans la pratique évaluative.

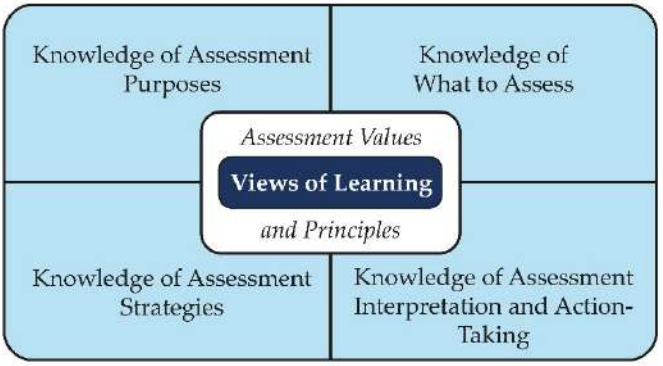


Figure 6 : Modèle de l'Assessment Literacy développé par Abell et Siegel (2011, p. 212)

Parmi les quatre axes prioritaires qu'elles définissent pour les études futures, elles mentionnent l'analyse des interactions entre la compétence évaluative, les pratiques effectives en classe et les résultats d'apprentissage. Leur proposition enrichit donc le cadre de Magnusson et

al. (1999) en y ajoutant deux catégories (finalités de l'évaluation et interprétation/prise de décision) et rejoint plus largement les considérations développées tant du côté du dernier modèle de consensus de Carlson et al. (2019) que du TALiP (Xu et Brown, 2016) quant à l'interaction des connaissances « mises en œuvre ».

IV. Discussion et perspectives

A travers les différents modèles théoriques présentés dans ce chapitre, nous avons tenté de mettre en exergue les articulations possibles entre deux champs de recherches majeurs, permettant de dépasser l'opposition entre didactique et évaluation (voir Bart, 2023 pour une tentative de synthèse). Quatre points de débat guideront cette dernière partie.

Tout d'abord, ce travail théorique nous conduit à discuter la question de la granularité des savoirs. En ce qui concerne les PCK pour les sciences, la question du grain des savoirs disciplinaires en jeu est posée dès 1999 par Veal et Makinster qui reconnaissent la nécessité de distinguer avec précision le rôle et le type de PCK pour pouvoir agir adéquatement sur le développement professionnel des enseignant·es de sciences. C'est dans ce cadre que les auteurs ont développé, en s'appuyant sur les manuels scolaires, une « taxonomie générale » des PCK permettant de classer les connaissances de diverses disciplines, matières et sujets scientifiques. D'après cette catégorisation, au premier plan figurent les PCK générales qui se réfèrent aux disciplines scolaires (histoire, maths, sciences, etc.). Au sein de ces disciplines s'inscrivent des PCK propres à un domaine spécifique (*Domain specific PCK*) (géologie, biologie, physique) et au sein de ces domaines interviennent des PCK liées à un objet de savoir particulier (*Topic specific PCK*) (pour la chimie : solubilité, oxydation, stœchiométrie). On retrouve le même type de réflexion du côté des mathématiques où une revue systématique récente (Grigaliùnienè et al., 2025) pose explicitement le questionnement du niveau de granularité approprié pour appréhender les PCK. Cette question de la granularité des savoirs pourrait dès lors constituer un premier levier pour étudier plus adéquatement les pratiques d'évaluation.

Le deuxième point de discussion porte sur la place de l'évaluation, qui n'est pas la même dans les modèles des deux disciplines. En sciences, l'*assessment* est nettement plus présente et se voit considérée dans un certain nombre de modèles en tant que composante même des PCK, bien qu'elle soit réintégrée dans la catégorie des connaissances de base dans les

derniers modèles de consensus (Carlson et al., 2019 ; Gess-Newsome et al., 2015). Dans les modélisations propres aux mathématiques, l'évaluation est généralement située dans les connaissances pédagogiques générales (Baumert et Kunter, 2013 ; Kaiser et Blömeke, 2013) et n'est pas formalisée comme une sous-composante explicite des PCK (Hill et al., 2008). Même si des problématiques liées à l'évaluation apparaissent clairement dans les descriptions plus fines de ces sous-dimensions (Baumert et Kunter, 2013), les modélisations investiguées ne semblent pas considérer les connaissances propres à l'évaluation en mathématiques comme un type de connaissances spécifiques qui vaudraient la peine d'être distinguées des autres types de connaissances spécifiques aux mathématiques. Aussi, en confrontant les différentes modélisations présentées, on peut considérer trois cas de figure : soit l'*assessment* est explicitée parmi les connaissances professionnelles générales, soit elle est explicitée parmi les sous-catégories de PCK, ou soit elle n'est pas explicitée comme une catégorie de PCK à part entière, même si elle est présente en filigrane. Lorsque l'évaluation n'est explicitement mentionnée que dans les connaissances générales et en filigrane dans les PCK, on peut craindre que les didacticien·nes n'accordent pas suffisamment de place à cette problématique et que les spécialistes de l'évaluation n'accordent pas suffisamment de place aux spécificités disciplinaires. Lorsque l'évaluation constitue explicitement une sous-dimension des PCK, on peut craindre que la focale sur les spécificités disciplinaires fasse perdre de vue l'apport des travaux majeurs dans le domaine de l'évaluation. Ces considérations nous conduisent alors à proposer d'investiguer dans les travaux futurs l'apport potentiel d'un nouveau concept : celui d'*Assessment Content Knowledge* (ACK[4]) qui, tout comme les PCK de Shulman (1986), se situerait à l'intersection entre les connaissances générales en évaluation et les connaissances disciplinaires. Plutôt que de chercher à élémenter encore davantage les catégories de connaissances, ce concept permettrait de questionner, dans une vision plus large, en quoi certaines pratiques évaluatives sont caractéristiques de telle ou telle discipline ou encore, de déterminer ce qui est constitutif de leur pertinence en lien avec les contenus enseignés.

Le troisième point de discussion est celui du lien, établi par plusieurs recherches entre PCK et/ou connaissances évaluatives, d'une part, et

[4] Ce concept a été évoqué pour la première fois par Annick Fagnant dans la discussion d'un symposium de l'ADMEE organisé en 2025 par Céline Lepareur et Raphaël Pasquini. C'est ce symposium, intitulé « *Quelles approches innovantes pour mettre (enfin) en synergie évaluation et didactiques* ? » qui est à l'origine du présent chapitre.

qualité des pratiques pédagogiques et/ou apprentissage des élèves d'autre part (voir Depaepe et al., 2013 ; Pastore, 2023 ; van Driel et al. 2014 pour des synthèses). Le modèle TALIP (Xu et Brown, 2016) ainsi que différentes modélisations proposées en mathématiques et en sciences, considèrent que les PCK, comme les différents types de connaissances évaluatives, font partie des *Knowledge base* des enseignant·es qu'il convient encore de mobiliser ou de transformer pour faire preuve de compétences professionnelles, développer une identité évaluative et/ou un enseignement de qualité. En sciences, le modèle de Gess-Newsome et al. (2015) est un exemple de ce processus : les connaissances sont un socle, soumis à des amplificateurs ou des filtres impliquant notamment les croyances et le contexte spécifique dans lequel l'activité d'enseignement s'exerce. En mathématiques, le modèle en cascade de Krauss et al. (2020), qui revisite en partie les travaux appuyés sur le modèle COACTIV (Kunter et al., 2013), propose une modélisation intéressante du cheminement complexe entre les PCK dont disposent les enseignant·es et l'apprentissage effectif des élèves. Pour affecter la qualité de l'enseignement, les PCK doivent notamment pouvoir être mobilisées en situation grâce à des compétences spécifiques permettant aux enseignant·es de percevoir les enjeux didactiques, de les interpréter et de prendre des décisions en fonction des situations rencontrées[5].

Enfin, si la question de savoir comment se développent les compétences évaluatives et les PCK a été évoquée, à des degrés divers, dans ce chapitre, notamment via le modèle TALIP (Xu et Brown, 2016) et les travaux de Abell et Siegel (2011) en sciences (voir aussi Lin et Rowland, 2016, pour une synthèse dans le champ des PCK en mathématiques), gageons que les futurs travaux qui s'y attelleront auront davantage pour but de montrer la richesse d'articuler certains cadres théoriques issus de l'évaluation et des didactiques plutôt que de les opposer. Il en va selon nous de la qualité du développement professionnel des enseignant·es et de l'amélioration des apprentissages des élèves.

[5] Ce modèle est proposé à titre illustratif en annexe 2.

V. Bibliographie

Abell, S. K. (2008). Twenty Years Later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, *30*(10), 1405-1416.

Abell, S. K. et Siegel, M. A. (2011). Assessment Literacy: What Science Teachers Need to Know and Be Able to Do. Dans D. Corrigan, J. Dillon et R. Gunstone (dir.), *The professional knowledge base of science teaching* (p. 205 221). Springer.

Allal, L. et Laveault, D. (2009). Assessment for Learning: Evaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation, 32*(2), 99–106.

Ball, D. L., Thames, M. H. et Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education, 59*(5), 389-407. https://doi.org/10.1177/0022487108324554

Bart, D. (2023). *Évaluation et didactique. Un dialogue critique*. Peter Lang.

Baumert, J. et Kunter, M. (2013). The COACTIV Model of Teachers' Professional Competence. Dans M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss et M. Neubrand (dir.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project* (p. 25-48). Springer.

Bennett, E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Blömeke, S. et Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM Mathematics Education, 44*(3), 223-247.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and practice, 30*(1), 3-12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., … Wilson, C. D. (2019). The Refined

Consensus Model of Pedagogical Content Knowledge in Science Education. Dans A. Hume, R. Cooper, et A. Borowski (dir.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (p. 77-94). Springer.

Chapman, O. (2015). Mathematics teachers' knowledge for teaching problem solving. *LUMAT, 3*, 19–36.

Cowie, B. et Moreland, J. (2015). Leveraging disciplinary practices to support students' active participation in formative assessment. *Assessment in Education: Principles, Policy & Practice, 22*(2), 247–264. https://doi.org/10.1080/0969594X.2015.1015960

Cross, D. (2025). Les Pedagogical Content Knowledge dans la tradition de recherche anglo-saxonne. Étude des fondements et des limites de quelques modèles en vue d'une classification. *Éducation et didactique* [En ligne], 19-3. https://doi.org/10.4000/153gq

Depaepe, F., Verschaffel, L. et Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12-25.

Depaepe, F., Verschaffel, L. et Star, J. (2020). Expertise in developing students' expertise in mathematics: Bridging teachers' professional knowledge and instructional quality. *ZDM Mathematics Education, 52*, 179-192. https://doi.org/10.1016/j.tate.2013.03.001

Depaepe, F. (2025, 25-29 août). *Improving PCK: insights from Teacher Education in History, Geography, and Biology in Five Countries* [Presentation]. EARLI 2025, Graz, Autriche.

Edwards, F. (2020). The effect of the lens of the teacher on summative assessment decision making: the role of amplifiers and filters. *The Curriculum Journal, 31*(3), 379-397. https://doi.org/10.1002/curj.4

Fagnant, A. (2023). *Les pratiques d'évaluation en classe : des compétences professionnelles pour soutenir l'apprentissage des élèves*. Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/08/Cnesco-CC-Eval_FAGNANT.pdf

Friedrichsen, P., Driel, J.H.V. et Abell, S.K. (2011), Taking a closer look at science teaching orientations. *Science Education, 95*(2), 358-376. https://doi.org/10.1002/sce.20428

Fulmer, G. W., Lee, I. C. H. et Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice, 22*, 475–494. https://doi.org/10.1080/2331186X.2016.1225380

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. Dans A. Berry, P. Friedrichsen et J. Loughran (dir.), *Re-examining Pedagogical Content Knowledge in Science Education* (p.28-42). Routledge. https://doi.org/10.4324/9781315735665-8

Grigaliūnienė, M., Lehtinen, E., Verschaffel, L. et Depaepe, F. (2025). Systematic Review of Research on Pedagogical Content Knowledge in Mathematics: Insights from a Topic-Specific Approach. *ZDM Mathematics Education*. https://doi.org/10.1007/s11858-025-01684-1

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press.

Hill, H. C., Ball, D. L. et Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372-400.

Kaiser, G. et Blömeke, S. (2013). Learning from the Eastern and the Western debate: the case of mathematics teacher education. *ZDM Mathematics Education, 45,* 7–19.

Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Studies in Science Education, 45*(2), 169–204. https://doi.org/10.1080/03057260903142285

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N. et Blum, W. (2020). Competence as a continuum in the COACTIV study : the "cascade model". *ZDM Mathematics Education 52*, 311–327. https://doi.org/10.1007/s11858-020-01151-z

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. et Neubrand, M. (2013). *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*. Springer.

Lin, F. L. et Rowland, T. (2016). Pre-Service and In-Service Mathematics Teachers' Knowledge and Professional Development. Dans A. Gutiérrez, G. C. Leder, et P. Boero (dir.) *The Second Handbook of Research on the Psychology of Mathematics Education* (p. 483–520). SensePublishers. https://doi.org/10.1007/978-94-6300-561-6_14

Loughran, J., Berry, A. et Mulhall, P. (2012). Portraying PCK. *Understanding and Developing Science Teachers' Pedagogical Content Knowledge*, (p. 15-23). SensePublishers. https://doi.org/10.1007/978-94-6091-821-6_3

Magnusson, S., Krajcik, J. et Borko, H. (1999). Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. Dans J. Gess-Newsome et N. G. Lederman (dir.), *Examining Pedagogical Content Knowledge* (vol. 6, p. 95-132). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47217-1_4

Park, S. et Oliver, J. S. (2008). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Research in Science Education, 38*(3), 261-284. https://doi.org/10.1007/s11165-007-9049-6

Park, S. et Suh, J. K. (2019). The PCK mapping approach to capturing the complexity of enacted PCK (EPCK) and pedagogical reasoning in science teaching. Dans A. Hume, R. Cooper et A. Borowski (dir.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (p. 185–197). Springer.

Pasquini, R. et Morales Villabona, F. (2022). Comment appréhender une littératie en évaluation pour documenter et comprendre le développement des compétences des enseignants ? *Revue canadienne de l'éducation, 45*(2), 512-554. https://doi.org/10.53967/cje-rce.v45i2.5171

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1217167

Pastore, S. et Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model, *Teaching and Teacher Education, 84*, 128-138, https://doi.org/10.1016/j.tate.2019.05.003

Pellegrino, J. W., Chudowsky, N. et Glaser, R. (dir.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academy.

Piñeiro, J. L., Chapman, O., Castro-Rodríguez, E. et Castro, E. (2021). Prospective Elementary Teachers' Pedagogical Knowledge for Mathematical Problem Solving. *Mathematics*, *9*(15), 1811. https://doi.org/10.3390/math9151811

Şen, M. (2023). Suggestions for the Analysis of Science Teachers' Pedagogical Content Knowledge Components and Their Interactions. *Research in Science Education, 53*, 1081–1095. https://doi.org/10.1007/s11165-023-10124-7

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher, 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22. https://people.ucsc.edu/~ktellez/shulman.pdf

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77(3*), 238–245.

van Driel, J., Verloop, N. et de Wos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *35*(6).

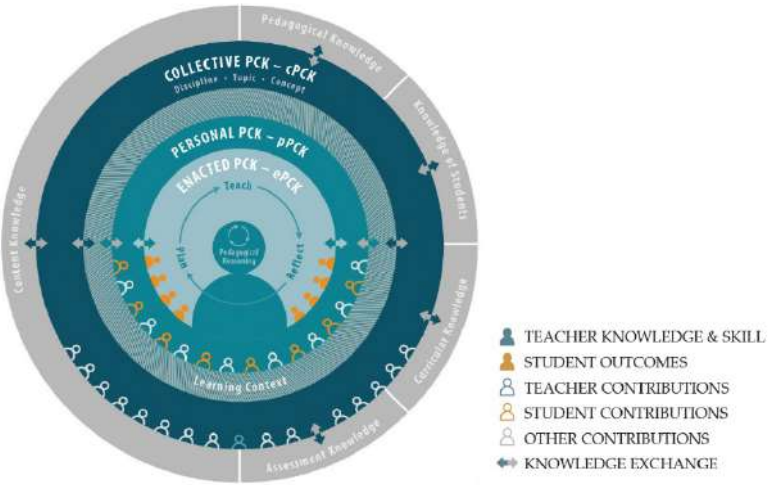
van Driel, J. H., Berry, A. et Meirink, J. (2014). Research on science teacher knowledge. Dans N. G. Lederman et S. K. Abell. (dir.), *Handbook of research on science education, volume II* (p. 862-884). https://doi.org/10.4324/9780203097267.ch42

Veal, W. et Makinster, J. (1999). Pedagogical Content Knowledge Taxonomies. *Electronic Journal of Science Education, 3*.

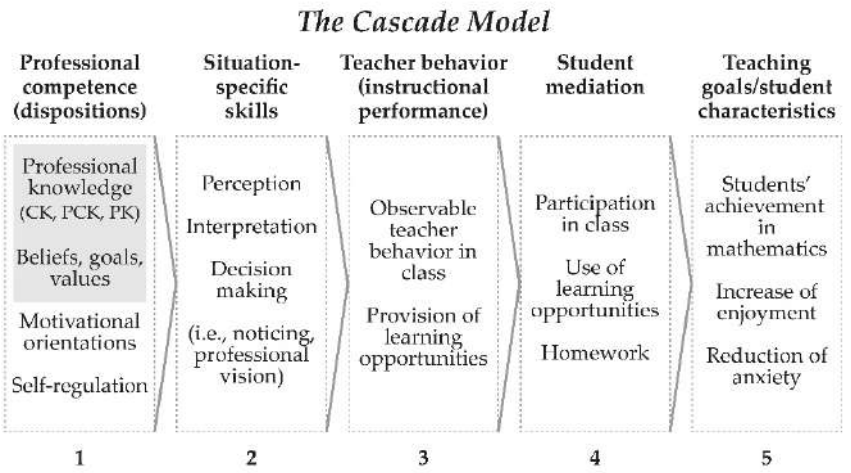
Willis, J., Adie, L. et Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher, 40*(2), 241-256. 10.1007/s13384-013-0089-9

Xu, Y. et Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010

VI. Annexes



Annexe 1 : Modèle issu de la 2ème conférence de consensus de 2017 (Carlson et al., 2019, p. 85)



Annexe 2 : Le modèle en cascade – Extrait de Kraus et al., 2020, p. 316

Chapitre 14. Évaluer la professionnalisation

Apport de 40 ans de publication de l'ADMEE

Christophe Gremion

I. Introduction

La question de la professionnalisation est apparue à la fin du XIXe siècle, dans les pays anglo-saxons puis européens, et fait référence tout d'abord à la « fabrication des professions » (Wittorski, 2014, p. 233). Ce mouvement permet à des individus exerçant la même activité de faire reconnaître leur rôle et accéder par là au droit d'exercer. Pour ce faire, ils mettent en avant des valeurs telles que l'autonomie ou la responsabilité individuelle. Plus tard, en Europe principalement, le vocabulaire de la professionnalisation a été utilisé par le monde de l'entreprise (responsabilisation, efficacité du travail, gain de performance…), ce qui est à mettre en lien avec l'organisation et le contrôle du travail dans le cadre du management. Enfin, dès la fin du XXe siècle, c'est le monde de la formation qui s'est emparé du concept. Une manière de répondre aux besoins de la société en préparant les individus à s'adapter à un environnement de travail qui change très rapidement (Jorro, 2014). Ainsi, la professionnalisation est autant un mouvement de reconnaissance des métiers vers plus d'autonomie, une attente institutionnelle vers plus d'efficacité et une volonté du monde de la formation vers une meilleure adaptation au contexte. Trois intentions parfois en opposition qui permettent de comprendre pourquoi ce concept fait souvent débat… Nous saisissons

l'occasion des 40 ans de l'ADMEE-Europe pour nous questionner sur la place de ce concept en sciences de l'éducation et de la formation, sur la signification que les auteurs lui donnent et - c'est le cœur du métier de cette association et l'intention de ce chapitre - sur les indicateurs mobilisés dans différents textes pour évaluer la professionnalisation. À cette fin, les articles publiés durant les 38 années de la revue « Mesure et évaluation en éducation » (MEE) et durant les 10 années de la revue « Évaluer. Journal international de recherche en éducation et formation » (e-JIREF) ainsi que les chapitres de l'ouvrage paru aux Presses de l'ADMEE ont été pris en compte.

II. Quelques éléments conceptuels

II.a. De quelle professionnalisation parle-t-on ?

Le trop bref historique proposé en introduction explique la polysémie de ce terme. Dans les trois contextes esquissés, la notion de professionnalisation se rapporte à l'amélioration des compétences d'un individu ou d'un groupe professionnel, à son engagement, à sa motivation et à sa reconnaissance, mais sa signification varie selon la perspective utilisée. Du point de vue des associations professionnelles ou des syndicats, on met en avant la professionnalisation axée sur la profession elle-même, se traduisant par des normes collectives telles que des règles communes ou des codes de déontologie visant à assurer la qualité et l'éthique du service fourni à la communauté ou visant à défendre les acquis et/ou intérêts de la corporation. On parle dans ce cas de professionnalisation-profession. Dans les entreprises privées ou institutions publiques, l'accent est placé sur l'efficacité professionnelle, en soulignant l'importance de l'employabilité, de l'adaptation rapide aux changements et de l'efficience. Dans ce cas, il est fait mention de l'idée de professionnalisation-efficacité du travail. Dans les établissements dédiés à la formation, le terme renvoie à ce que l'on appelle professionnalisation-formation. L'objectif est alors de renforcer les compétences pour permettre à l'individu de prendre des décisions autonomes, réfléchies et responsables face à des situations imprévues.

Ainsi, ces trois définitions émergent de contextes distincts : l'identité et la réputation de la corporation vis-à-vis de la société, le marché du travail avec son besoin d'efficacité et enfin la formation et la prise de décisions autonomes et responsables (Gremion, 2025, p. 32 ; Wittorski, 2014, p. 235).

Dans le domaine de la formation, professionnalisation et professionnalité émergente sont deux concepts souvent utilisés comme synonymes. Il faut reconnaître que, si durant le XXe siècle, il était possible de différencier facilement la professionnalisation des métiers avec la professionnalité (émergente) des personnes, le premier terme ne concernant que l'activité du collectif alors que le deuxième s'adresse aux acteurs ; ce n'est plus le cas actuellement. En effet, une recherche rapide sur la plateforme Cainr.info donne des résultats tels que professionnalisation des ingénieurs, des élus, des formateurs, mais aussi de l'enseignement, des associations, ou encore des formations, des études… Tout et tout le monde semblent enclins à se professionnaliser.

S'appuyant sur les travaux de De Ketele et Jorro (2013), Graber et Haberey-Knuessi (2017, p. 69) donnent une définition intéressante de la professionnalité : « La vision de la professionnalité est plus spécifique et correspond à une photographie instantanée de la professionnalisation. La professionnalité est ainsi plus statique, elle réunit des savoirs d'action, de stratégie, de méthode et des savoirs éthiques ». Ces autrices évoquent à ce sujet « l'idée de qualité de l'activité professionnelle, de consistance dans l'expertise du professionnel ». Pour la définition de la professionnalisation, elles évoquent le terme de processus, disant qu'il « est donc dynamique et convoque à la fois l'engagement, la motivation, la reconnaissance et la professionnalité émergente » (2017, p. 69). Ainsi, derrière le terme de professionnalisation se retrouvent l'appropriation du genre professionnel (la professionnalité), mais aussi le souci de performance, le renforcement du sens des responsabilités et l'augmentation de l'efficacité du travail. Ce sont donc bien les trois types de professionnalisation (-formation, -profession, -efficacité du travail) qui sont convoqués dans ce cas. Nous partageons ces définitions et la différence qui peut être faite entre professionnalisation et professionnalité, raison pour laquelle notre attention ne se portera que sur l'utilisation du premier.

II.b. Professionnalisation de qui ou de quoi ?

Dans l'étude des processus de professionnalisation, trois tendances se succèdent (Wittorski, 2014, p. 233). La première, anglo-saxonne, s'intéresse à la manière dont les personnes exerçant une même activité s'organisent et donnent à voir leurs règles communes pour être reconnues professionnellement. Une deuxième, en réaction critique face à la première, s'intéresse plus à l'individu qu'au groupe et voit la professionnalisation « comme étant au service d'une nouvelle mobilisation

des salariés dans des contextes de travail plus flexibles, faisant davantage appel aux ressources subjectives des personnes » (p. 234). Enfin, après le groupe et l'individu, c'est également la formation qui « va être pénétrée progressivement par les discours sur la professionnalisation (Maubant et Gremion, 2022, p. 6), modifiant ses ingénieries pour mieux articuler travail et formation dans des architectures que l'on nomme duales ou en alternance. L'analyse que nous avons faite des textes retenus a donc visé à observer à qui ou à quoi s'intéresse la professionnalisation : l'individu, le groupe métier ou la formation.

II.c. Professionnalisation pour quoi ?

Une première raison qui vient parfois à l'esprit est de voir la professionnalisation comme mouvement s'opposant à l'amateurisme ou au bénévolat (Perrenoud, 1993). Une définition qui n'est généralement pas utilisée dans le monde de la formation, mais qui s'emploie encore régulièrement, par exemple dans le domaine militaire : là, professionnalisation représente la transformation de l'armée de milice en une armée professionnelle.

Les auteurs utilisent aussi le terme de professionnalisation pour signifier une amélioration des gestes professionnels. C'est ce que l'on retrouve en général derrière la notion de performance comme dans l'exemple de formations professionnelles qui utilisent « le suivi oculaire comme un outil innovant au service de la professionnalisation des enseignants » (Duvivier, 2025, p. 1). Cette définition est à mettre fortement en lien avec celle de professionnalité évoquée ci-dessus.

Mais, dans la littérature, les deux éléments les plus souvent évoqués comme raisons du besoin de ce mouvement de professionnalisation sont 1) la nécessité de développer les compétences, ces savoir-agir en situation, cette capacité à faire des choix face à l'imprévu ou à s'adapter à l'imprescriptibilité et 2) l'intention de développer l'autonomie et la responsabilisation (Perrenoud, 1993 ; Wittorski, 2014), dimensions qui s'articulent « dans une dynamique favorable à l'émancipation » (Moukaddam, 2024, p. 9). L'analyse que nous avons faite des textes retenus a donc visé également à observer les raisons de cette recherche de professionnalisation : pour quitter l'amateurisme, pour développer les gestes professionnels, pour permettre les choix en situation inconnue ou pour favoriser l'émancipation (englobant autonomie et responsabilisation).

II.d. Quels indicateurs de professionnalisation ?

Pour observer et catégoriser les indicateurs retenus par les auteurs, nous nous appuyons sur les cinq domaines de professionnalisation mis en évidence dans différents travaux (Gremion et al., 2021 ; Paseka et al., 2010 ; Piot, 2009) :

Réflexivité et discours sur son action : concerne le degré de cohérence et de pertinence du discours sur l'analyse et la régulation de son activité, la capacité à autoévaluer son action en tenant compte d'un genre professionnel reconnu et attendu, la qualité de l'anticipation des potentielles difficultés issues d'une situation, la régulation de son activité en fonction de son autoévaluation ainsi que la pertinence de la problématisation d'une situation.

Conscience de son identité professionnelle : ce domaine s'intéresse à l'expression du sentiment d'appartenance à un collectif, à la mise en évidence des écarts entre son propre style professionnel et un genre professionnel reconnu, à la prise des décisions d'action en toute autonomie, à l'aptitude de définir son rôle et sa responsabilité vis-à-vis des apprentis, de l'institution et de la société et enfin à l'habileté à justifier ses actes en s'appuyant sur des valeurs personnelles.

Coopération et collégialité : sont incluses dans ce domaine l'expression de tension et/ou d'adhésion à un genre professionnel reconnu par le collectif, l'identification des différences entre son style professionnel et un genre professionnel attendu par un collectif d'appartenance, la justification de ces différences en s'appuyant sur des valeurs tant personnelles que socialement reconnues et enfin la capacité à modifier son style professionnel pour être plus en phase avec un genre attendu par le collectif.

Adaptation à la complexité et à la singularité de chaque situation : dans ce domaine, nous observons l'aptitude à justifier ses prises de décision en fonction de la singularité de la situation et de ses valeurs personnelles, la capacité de tenir compte des besoins spécifiques inhérents à chaque situation, l'analyse a priori des situations nouvelles ainsi que la facilité à trouver de nouvelles solutions pour gérer des situations inattendues.

Niveau de maîtrise personnelle : ce domaine comprend notamment l'expression de la conscience de compétence et du besoin de formation, la reconnaissance de ses forces et ses faiblesses dans l'exercice de sa profession, l'identification des domaines nécessitant un complément de

formation et le souci de faire évoluer ses compétences en fonction des modifications de l'environnement, du public, des technologies, des innovations éducatives...

III. Méthodologie

Pour mener cette recherche, nous nous sommes intéressés aux trois outils de publication de l'ADMEE, soit la revue Mesure et Evaluation en Education (MEE), Evaluer. Journal international de recherche en éducation et formation (e-JIREF) et les Presses de l'ADMEE.

Pour MEE, tous les articles de 1978 à 1995 ont été consultés en PDF, aucun article n'a été retenu puisqu'aucun n'abordait dans son titre ou son résumé la question de la professionnalisation. Pour les numéros de 1995 à 2025, la requête « professionnalisation site:https://www.erudit.org/fr/revues/mee/ » a été utilisée pour effectuer une première sélection de 157 références. Dans celles-ci n'ont pas été retenues les notices, ni les occurrences qui renvoient à un sommaire de numéro ou une entrée bibliographique. Dans les articles restants n'ont été retenus que les articles 1) qui donnent à voir une définition de la professionnalisation, 2) ou qui identifient (ou permettent de comprendre) des dimensions de la professionnalisation, 3) ou qui rendent visibles des indicateurs de professionnalisation. Au total pour MEE : 15 articles sont retenus. Pour e-JIREF, la recherche s'appuie sur la même utilisation de la requête « professionnalisation site:https://journal.admee.org » pour une première sélection de 31 articles. Le même travail de sélection que pour MEE a eu lieu, ce qui a permis de retenir au final cinq articles. Pour les Presses de l'ADMEE, nous avons pris l'ouvrage actuellement publié et avons procédé selon la même démarche de recherche et de sélection. Un chapitre a pu être retenu sur la base des critères évoqués ci-dessus. Tous les articles ou chapitres pris en compte se retrouvent dans la bibliographie, identifiable par l'astérisque qui précède le nom du premier auteur.

Une analyse catégorielle des 21 articles a eu lieu ensuite, avec une focalisation sur les passages utilisant le mot « professionnalisation ». Les catégories retenues dans le cadre conceptuel ont été utilisées, soit : l'objet de la professionnalisation (le métier, la formation ou les personnes), la base de définition de la professionnalisation (quitter le statut d'amateur, améliorer le geste professionnel, développer la capacité à faire des choix ou développer l'autonomie et la responsabilisation à des fins d'émancipation) et les indicateurs de professionnalisation (réflexivité et

discours sur son action, conscience de son identité professionnelle, coopération et collégialité, adaptation à la complexité et à la singularité de chaque situation et niveau de maîtrise personnelle). Les résultats de cette analyse sont présentés dans la section suivante.

Naturellement, ce travail comporte de nombreuses limites, nous relèverons ici les deux principales : la première réside dans le fait que nous n'avons consulté que les articles et chapitres issus des outils de publication de l'ADMEE, c'était notre choix de départ pour contribuer à cet ouvrage des 40 ans de l'ADMEE-Europe. Nous aurions eu évidemment un choix beaucoup plus large en nous intéressant à d'autres revues ou maisons d'éditions, certaines ayant fait de la professionnalisation le centre de leur ligne éditoriale. La seconde limite vient du choix d'analyser les textes utilisant littéralement le mot professionnalisation et d'écarter les articles traitant de la professionnalité ou du développement professionnel. Nous l'avons vu, les deux termes sont très proches et il est possible que nous ayons pu élargir notre échantillon en modifiant ce critère de choix. Nous avions cependant la crainte de ne pas respecter le choix des auteurs, l'emploi d'un terme ou de l'autre n'étant certainement pas anodin.

IV. Résultats

IV.a. Les objets concernés par la professionnalisation

Les trois objets présentés dans le cadre conceptuel sont observés, soit la professionnalisation du métier, la professionnalisation de la formation (développement de formations professionnalisantes) et la professionnalisation des personnes.

Entre 1978 et 2007, neuf textes abordent la question de la professionnalisation, dont quatre font référence à la professionnalisation des métiers. C'est le cas de Perrenoud (1996, p. 74) : « une réforme peut être un moyen de créer une dynamique, de restaurer un climat, de donner une image positive, de dépasser des conflits ou des blocages, de favoriser une évolution vers la décentralisation, la professionnalisation, mais aussi la marginalisation ou la neutralisation de certaines catégories d'acteurs ». Cette notion de « catégories d'acteurs » renvoie au collectif exerçant le même métier. Le même objet d'attention est mis en évidence dans le texte de Garon et Théorêt (2005, p. 2) qui s'intéressent à « […] la

professionnalisation de l'enseignement ». Ici, « l'enseignement » renvoie à l'ensemble des acteurs qui exerce ce même métier.

Dans leur article, Briquet-Duhazé et Buhot (2007, p. 9) observent la professionnalisation de l'enseignement (en tant que métier) en la différenciant de la notion de professionnalité « Pour Aballea (1992), si le champ de l'enseignement ne remplit pas toutes les conditions pour faire du métier d'enseignant une profession, il est cependant possible de viser le développement d'une professionnalité, système complexe et composite d'expertise, reconnu, scientifiquement validé, et encadré par un système de références, de valeurs et de normes. »

Enfin, Vial, dans son texte de 1998, parle de la professionnalisation du métier « alors les étudiants devraient pouvoir communiquer à propos des orientations, des visées, des valeurs éthiques et politiques dans le cadre d'une profession, l'expression du sens de leurs projets, de leurs postures devant la professionnalisation, dépassant le discours organisationnel sur la formation » (1998, p. 65) mais il évoque également le projet de professionnalisation des individus « L'évolution, le développement, l'enrichissement des personnes, de leur processus d'apprentissage et de leur projet de professionnalisation ; leur socialisation, en somme, ne peut-elle être parlée ? » (p. 58).

Entre 2008 et le printemps 2025, 12 textes traitent de la professionnalisation, dont seulement deux en s'intéressant à la professionnalisation des métiers : Fernandez et Roy (2010, p. 2) comparent l'évolution des normes d'éthique en testing aux États-Unis avec des normes présentes dans d'autres professions (médecine, génie, administration, etc.) et Moukaddam (2024, p. 9) aborde clairement la professionnalisation de l'enseignement en tant que métier ou collectif d'acteurs exerçant la même activité.

Pour les autres textes, deux s'intéressent à la formation à des fins de professionnalisation et sept, la grande majorité, annoncent parler de la professionnalisation des personnes. C'est le cas de Figari (2008, p. 81) qui fait état de la professionnalisation des enseignants, de Manier et Abernot (2016, p. 56) qui parle d'un « connais-toi toi-même » qui renvoie clairement aux personnes, de Gremion et Coen (2016, p. 12) qui s'intéressent à la professionnalisation des praticiens « Un constat qui impose aux praticiens d'être en mesure de constamment résoudre de nouveaux problèmes, d'inventer de nouvelles solutions, de s'adapter à de nouvelles situations » ou encore le Born (2023, p. 106) « un levier pour la professionnalisation des sapeurs-pompiers volontaires ».

Enfin, un texte se positionne dans la dimension de formation professionnalisante (Betbeder Laüque, 2019, paragr. 1) « Par professionnalisation, il faut entendre "une tendance vers l'ajustement aux besoins du marché du travail reflétée dans le choix des programmes de formation fait par les étudiants" (Crespo, 2013, p. 4). Pour ce qui est du contexte européen, le processus de Bologne semble avoir impulsé la création de programmes professionnalisants à l'instar des licences professionnelles en France (1999) et des foundation degrees en Angleterre (2001). » Ce passage met en évidence l'articulation théorie-pratique et la place de l'alternance dans les formations dites « à visées professionnalisantes ».

Ainsi, jusqu'en 2007, la réflexion sur la professionnalisation concerne principalement la professionnalisation des métiers, de la formation et des individus, en soulignant l'importance de développer une expertise reconnue et une identité collective. À partir de 2008, l'accent se porte davantage sur la professionnalisation des personnes et des formations, intégrant notamment l'ajustement aux besoins du marché et le rôle de l'alternance dans la formation professionnalisante.

IV.b. Les définitions de la professionnalisation

Au début de ce travail d'analyse des textes de e-JIREF, MEE et des Presses, nous avions pensé trouver quatre types de définitions : celle qui définit la professionnalisation comme le fait de quitter le statut d'amateur, celle qui insiste sur l'amélioration du geste professionnel, celle s'intéressant à la capacité de faire des choix et enfin celle se centrant sur le développement de l'autonomie et de la responsabilisation à des fins d'émancipation. Sur les 21 textes retenus, seuls six nous ont permis d'identifier la définition utilisée par les auteurs. Ce constat semble corroborer les propos de Maubant et al. (2010) qui suggèrent que ce concept est devenu un allant de soi qui n'est plus questionné, du moins dans les textes issus de notre sélection.

Un premier texte semble prendre appui sur l'amélioration du geste et la capacité à faire des choix en développant ses compétences. Il s'agit de celui de Perrenoud (1996, p. 89) : « formation initiale et continue des enseignants […] doit favoriser la professionnalisation et développer les compétences requises à tous les niveaux précédents ».

Les cinq autres textes qui donnent à voir une définition présentent le concept comme une recherche d'autonomie et responsabilisation s'articulant au service de l'émancipation des personnes ou des collectifs.

Ainsi, Vial (1998, p. 57) met en lien la dimension de l'autonomie avec celle d'émancipation, que l'on peut comprendre comme visée du processus de professionnalisation lorsqu'il fait la critique du clientélisme en formation : « Enfin, quittons les modèles de l'évaluation pour rappeler que, dès l'antiquité, le mot de « client » ne peut rien connoter du côté de l'autonomie mais bien au contraire implique nécessairement un rapport de dépendance. Considérer le formé comme un client à satisfaire, c'est le manipuler suffisamment pour le fidéliser. Il faut ne rien connaître des théories de l'engagement pour croire que le clientélisme serait une marque d'émancipation. » Jorro, dans son article de 2004, utilise également les notions d'autonomie dans sa définition : « La recherche dans le champ de la professionnalisation des enseignants et des formateurs valorise la figure du praticien réfléchi, gagnant une autonomie de pensée et d'action (Anadon, 1999). » (2004, p. 43). Cette même autonomisation est également évoquée par Rouiller et Pillonel (2004, p. 52) comme condition nécessaire à la professionnalisation. Gremion et Coen (2016, p. 12) ajoutent à l'autonomie le besoin de responsabilisation pour guider la prise de décision. Enfin, Moukaddam (2024, p. 9) rapproche les notions de professionnalisation et de professionnalité émergente au service de l'émancipation des acteurs.

Nous voyons donc que la plupart des textes analysés expliquent la professionnalisation comme un processus visant l'autonomie, la responsabilisation et l'émancipation des individus ou des collectifs, plutôt que simplement la maîtrise technique ou le changement de statut. Seuls quelques-uns insistent sur le développement des compétences et la capacité à faire des choix comme fondements de cette démarche.

IV.c. Les indicateurs mobilisés pour observer la professionnalisation

Cinq indicateurs ont été retenus pour l'analyse des apports des textes, tous retrouvés : réflexivité et discours sur son action (11 textes), conscience de son identité professionnelle (10 textes), coopération et collégialité (4 textes), adaptation à la complexité et à la singularité de chaque situation (8 textes) et niveau de maîtrise personnelle (6 textes). Notre lecture des différents textes n'a pas fait apparaître d'autres indicateurs.

Réflexivité et discours sur son action :

Cet indicateur, cité dans 11 textes, est, avec la notion d'identité professionnelle, celui qui fait le plus consensus. Perrenoud utilise la réflexion sur sa propre pratique confrontée aux savoirs issus de la recherche comme signe de professionnalisation (1996, p. 90) : « Dans le domaine de l'enseignement, c'est une manière d'être parmi d'autres, qu'on justifiera volontiers en disant que « tant que les chercheurs ne se mettent pas d'accord sur ces questions, inutile d'aller perdre son temps en lectures stériles ». Une telle justification signifie tout simplement que la professionnalisation est encore faible ou qu'elle se développe à l'écart d'une culture scientifique de base, comme simple identité corporative sans savoirs de référence ». D'autres auteurs mettent en avant la réflexivité comme réponse à l'imprévisibilité des situations rencontrées par les professionnels : « Pour permettre leur professionnalisation, pour les rendre suffisamment autonomes et responsables pour gérer toute l'incertitude planant entre travail prescrit et travail réel, les futurs enseignants ont besoin de bonnes compétences afin de réfléchir « sur » et « dans » leur action. » (Gremion et Coen, 2016, p. 12).

Conscience de son identité professionnelle :

Indicateur fréquemment avancé également (10 textes), l'identité professionnelle est mise en lien avec le genre professionnel (qui semble en être le socle) et la responsabilité qu'a chaque acteur de s'y conformer. Ainsi, Vaudroz et Berger parlent d'« une multitude de compétences, de capacités et d'aptitudes portant notamment sur les savoirs spécifiques au métier, sur les représentations de normes collectives sur lesquelles se construit l'identité professionnelle ». (2018, p. 89). D'autres auteurs insistent sur le besoin de comprendre la construction de l'identité professionnelle chez les novices : « Il s'agit alors d'appréhender, d'une part, les mécanismes de construction de l'identité professionnelle, de construction et de développement des compétences professionnelles chez les futurs enseignants et, d'autre part, les éléments qui constituent ce que l'on a appelé plus haut la professionnalité des futurs maîtres. » (Briquet-Duhazé et Buhot, 2007, p. 10).

Ce lien entre genre professionnel et identité professionnelle est nuancé par ce que l'on peut comprendre comme une recherche de son propre style professionnel (De Ketele, 2017). C'est ce que proposent Manier et Abernot : « L'idée n'est pas d'uniformiser les attitudes, ni d'obtenir des clones au fonctionnement identique, mais bien de tendre à un connais-toi toi-même toujours en question, en quête d'une professionnalisation qui tient en même temps du jugement subjectif et de l'exigence institutionnelle. »

(2016, p. 56). Enfin, Vial propose de voir la construction identitaire comme un mouvement perpétuel, ou un processus sans fin : « La construction de l'identité professionnelle, inachevable comme tout processus du sujet au travail, et constitutive de la professionnalisation, passe par la mise en circulation d'images emblématiques du bon professionnel : des effigies » (Vial, 2006, p. 92).

Coopération et collégialité :

Moins fréquemment citée (4 textes), la dimension collective de la professionnalisation se perçoit principalement derrière l'idée de socialisation (Vial, 1998, p. 58). Born nous dit, en parlant de la formation initiale, qu'« elle amorce en premier lieu la construction de l'identité professionnelle de l'individu, en l'immergeant dans le monde des sapeurs-pompiers et en lui offrant une première image de l'activité et de la figure du sapeur-pompier. L'apprenant s'inscrit en ce sens dans un processus identitaire auquel le formateur participe activement, en partageant notamment avec lui des valeurs et des normes propres à son travail et en lui offrant une immersion dans la culture professionnelle, qui vont petit à petit lui permettre de s'identifier au rôle qui est attendu de lui (Hughes, 1958). Les processus de socialisation professionnelle participent ainsi à la construction de l'identité professionnelle » (Born, 2023, p. 106). Mais si nous avons mis en évidence l'opposition et la complémentarité entre genre et style dans le paragraphe précédent, il en est de même de la dimension collective qui entre en concurrence avec le besoin de différenciation et d'individualisation : « Bien qu'accordant une place différente au sujet apprenant, ces fonctions s'articulent dans un processus de professionnalisation qui nécessite notamment individualisation, socialisation et autonomisation. » (Rouiller et Pillonel, 2004, p. 52).

Adaptation à la complexité et à la singularité de chaque situation :

L'adaptation à la complexité des situations apparaît dans 8 textes de notre corpus. Jorro (2004, p. 35) la rend visible derrière la notion de « pensée critique, [vue] comme facteur de professionnalisation, [qui] est identifiée par les praticiens lorsque les conditions de réassurance, de non-jugement, de réflexivité sont valorisées ». Figari, pour sa part, insiste sur la notion de situation pour justifier le besoin d'adaptation : « la mise en œuvre des compétences se fait dans des situations interactives, contextualisées, particulières […]. On voit apparaître, ici, la notion de « situation » […] qui, lorsqu'elle ne définit pas un contexte de déterminisme quelconque (sociologique, scolaire, etc.), exprime une organisation intentionnelle des conditions de travail ou d'apprentissage. » (Figari, 2008, p. 81). Pour Rouiller et Pillonel, si chaque situation est singulière et demande

l'adaptation à cette singularité dans l'action, il ne faut pas oublier que chaque individu, en formation ou au travail, est également singulier, d'où le besoin d'individualisation dont ils parlent (2004, p. 52). Afin de se préparer à cette imprévisibilité des situations, Neuhaus rappelle le besoin d'adaptabilité par la recherche du développement de la réflexion anticipatrice ou réflexion pour l'action. « A des fins de professionnalisation, Barbier et Wittorski (2015, p. 10) proposent de mettre en place des moments de réflexion rétrospective ou anticipatrice à propos de l'activité (analyse de pratiques, groupes de pairs par exemple) » (Neuhaus, 2022, p. 87). Enfin, Moukaddam insiste sur l'autonomie qui est une condition à l'adaptabilité face à la singularité des situations rencontrées : « En matière d'éthique, une telle autonomie est fondatrice pour la résolution des problèmes et la construction d'un rapport lucide à soi, à l'autre et au monde. » (2024, p. 7).

Niveau de maîtrise personnelle :

Ce dernier indicateur est présent dans 6 textes. Les auteurs relèvent l'importance de la place de la connaissance dans le processus de professionnalisation, processus qui « passe nécessairement par l'édification d'une base de connaissances (Gauthier et al., 1997), entraînera un rapprochement entre les enseignants et la recherche » (Garon et Théorêt, 2005, p. 2). Ce rapprochement entre praticien et recherche, ou plus communément entre théorie et pratique, est très présent lorsque les auteurs évoquent le niveau de maîtrise personnelle. Grangeat identifie ce thème comme l'un des trois objets principaux des thèses traitant de l'évaluation et soutenues au début des années 2000, « le transfert des acquis de la formation dans les pratiques, la cohérence dans les formations par alternance » au service de la professionnalisation (2002, p. 35). La recherche de liens théorie-pratique est également mise en évidence comme finalité des dispositifs de formation en alternance par de nombreux auteurs du panel (Anderführen Fiaux, 2004 ; Briquet-Duhazé et Buhot, 2007 ; Grangeat, 2002 ; Luisoni et Monnard, 2018). Les deux extraits ci-dessous montrent que l'alternance est pensée comme confrontation de différents points de vue au service d'une formation professionnalisante. « Les étudiants vivent donc dans deux milieux de formation, milieux dont les cultures sont fort différentes, bien que concourant, pour eux, au même but : la professionnalisation » (Anderführen Fiaux, 2004, p. 70). « Ce savoir agir stratégique est issu d'un enrichissement mutuel entre des savoirs pratiques constitués dans l'expérience et des savoirs théoriques formels et codifiés. » (Luisoni et Monnard, 2018, p. 83)

Ce lien théorie-pratique fort offert par les situations d'alternance fait dire à Maubant que « l'alternance constitue dans un certain nombre de cas un marqueur de l'objectif de professionnalisation des parcours de formation. Dès lors, le concept de professionnalisation se précise au fur et à mesure de son attelage au concept d'alternance. » (Maubant, 2020, p. 203)

Enfin, ce dernier extrait met en évidence les liens qui existent entre certains de ces indicateurs : « Trois dimensions semblent récurrentes lorsqu'il est question de professionnalisation : 1) le développement des compétences professionnelles, 2) l'appropriation de la culture professionnelle et 3) la construction d'une identité professionnelle » (Bélisle et al., 2022, paragr. 5). Ces trois dimensions fréquemment citées mettent en évidence la complexité de la notion de professionnalisation, liant à travers les pratiques tant le genre professionnel que le style personnel.

Ainsi, cinq indicateurs principaux (réflexivité, conscience de l'identité professionnelle, coopération, adaptation à la complexité et maîtrise personnelle) sont utilisés pour observer la professionnalisation, les plus présents étant la réflexivité, l'identité professionnelle et la capacité d'adaptation face à des situations singulières, ce qui illustre la complexité et la multidimensionnalité du processus.

V. Discussion et conclusions

L'analyse des 21 textes retenus nous permet quelques constats et questionnements.

Tout d'abord au niveau de l'objet de professionnalisation, au fil des années, la professionnalisation des métiers laisse petit à petit la place à celle des personnes. Est-ce une diminution de l'importance du collectif dans les travaux de recherche ou dans la société en général ? Ou simplement une place redonnée à l'individu ? Ou les travaux présentés émanent-ils principalement de formateurs-chercheurs qui, par leur rôle de formateur, prennent prioritairement l'effet de leur action sur les formés comme indicateur d'efficacité des formations ?

Au niveau des indicateurs relevés, éléments essentiels aux méthodologies d'évaluation, réflexivité, identité et adaptation sont les indicateurs les plus souvent mobilisés. Le niveau de maîtrise ne vient qu'en 4è ordre d'importance, avant la dimension collective. Pas d'autres indicateurs n'ont été mis en évidence dans les textes. Comment interpréter le fait que la dimension collective soit, là aussi, peu présente ? A-t-elle moins

d'importance dans les cursus de formation professionnelle ? Ou entre-t-elle trop en opposition avec la dimension individualiste de l'évaluation en contexte de formation ? Une autre réflexion pourrait être menée au sujet de la maîtrise personnelle (compétence métier) qui est moins présente que la réflexivité, l'adaptabilité et la construction identitaire, trois compétences souvent qualifiées de transversales. Les compétences transversales seraient-elles plus importantes que les compétences métiers dans le monde incertain dans lequel nous vivons ? Sont-elles plus utiles à l'employé de demain qui devra exercer un métier qui n'existe pas encore ?

Autre point d'attention, nous ne relevons que peu de définitions de la professionnalisation, avec un consensus autour de la professionnalisation-formation qui cherche l'autonomie, la responsabilité, voire l'émancipation des personnes aux injonctions souvent paradoxales. Est-ce une direction prise assez naturellement par les auteurs, de par leur fonction dans les systèmes de formation ? Ou une sensibilité du monde de la formation, voire de la société, qui valorise le travail dans les règles de l'art (professionnalisation-formation) et se méfie des normes que se fixent les travailleurs entre eux (professionnalisation-profession) ? Et qui se méfie tout autant des normes dictées par la recherche d'efficacité et de rentabilité (professionnalisation-efficacité du travail) ? Quoi qu'il en soit, force est de constater que les deux autres formes de professionnalisation (-efficacité du travail et -profession) semblent ainsi absentes des textes analysés. Y a-t-il une professionnalisation plus souhaitable que les autres ? Et en les mettant ainsi en retrait, ne court-on pas le risque d'une dialectique stérile, une dialectique qui mène à un bras de fer, une lutte de pouvoir entre travailleurs – formateurs – employeurs ? Un signe peut-être, dans le corpus analysé, seul un texte fait mention de la déprofessionnalisation (André, 2015, p. 68) « Et les délégitimations de l'objet par les formateurs de terrain présentes dans nos données (formulation « absconse » ; « prise de tête » ; « jargon professionnel » …) peuvent être comprises comme une résistance à ce pouvoir et une lutte contre une désappropriation du métier et les risques d'une déprofessionnalisation. » Il est intéressant de voir que, dans cet extrait, la parole est donnée aux formateurs de terrain. Ces derniers ne se reconnaissant pas (identitairement parlant) derrière le « jargon » universitaire. Une résistance à ce pouvoir du savoir universitaire, signe d'un déséquilibre entre la valeur des savoirs de terrain et des savoirs portés par les centres de formation ?

Toutes ces questions pourraient trouver leur place dans de futures recherches. Mais en guise de conclusion, nous nous permettons d'en ajouter encore une : les personnes employées souhaitent-elles, dans leur majorité, plus d'autonomie et de responsabilité ? Souhaitent-elles

effectivement se professionnaliser ou cette intention n'émane-t-elle que du politique, des employeurs, des formateurs ou des collectifs de travailleurs (syndicats ou corporations) ?

De là, une nouvelle série de questions s'impose. Est-ce lorsque les personnes employées souhaitent réellement s'émanciper de l'autorité, de la formation ou de la co rporation que le métier se professionnalise ? Mais s'émanciper, c'est autant accepter que s'opposer consciemment à la prescription. Dès lors, la désobéissance serait-elle facteur de professionnalisation ? Pourrait-elle être vue comme un sixième indicateur de professionnalisation des acteurs ?

VI. Bibliographie

*Anderführen Fiaux, S. (2004). Expériences de l'auto-évaluation d'étudiants en soins infirmiers : reflets d'une étude qualitative. *Revue Mesure et évaluation en éducation*, *27*(2), 69-89.

*André, B. (2015). L'usage du référentiel de compétences par les formateurs de terrain pour évaluer les étudiants en stages1. *e-JIREF*, *1*(3), 57-72. https://doi.org/10.48782/g4txvw69

*Bélisle, M., Heilporn, G., Lavoie, P., Lakhal, S., Lechasseur, K., Fernandez, N., Caty, M.-E. et Chichekian, T. (2022). Élaboration et validation d'une échelle de mesure de la professionnalisation des étudiants et des étudiantes universitaires en sciences de la santé. *Mesure et évaluation en éducation*, *45*(2), 69-105. https://doi.org/10.7202/1105562ar

*Betbeder Laüque, A.-L. (2019). Élaboration du Gabarit d'évaluation de l'environnement des programmes : le cas d'une unité universitaire de formation continue au Québec. *Mesure et évaluation en éducation*, *42*(2), 109-129. https://doi.org/10.7202/1071517ar

*Born, P. (2023). La sélection informelle en formation initiale de sapeur-pompier volontaire. *e-JIREF*, *9*(1), 103-121. https://doi.org/10.48782/e-jiref-9-1-103

*Briquet-Duhazé, S. et Buhot, É. (2007). Les rapports de visite des formateurs sur le stage en responsabilité des professeurs des écoles stagiaires : une diversité révélatrice de conceptions du

savoir-enseigner. *Mesure et évaluation en éducation*, *30*(1), 5-29. https://doi.org/10.7202/1086170ar

De Ketele, J.-M. (2017). Concilier norme et singularités : comment utiliser nos marges de liberté ? *Administration & Éducation*, *155*(3), 85-102. https://doi.org/10.3917/admed.155.0085

Duvivier, V. (2025). Le suivi oculaire comme outil d'analyse des gestes professionnels : application au projet ARC Sim'Pro. *Enseignement & Apprentissages*, (13), 1-19.

Fernandez, A.-S. et Le Roy, F. (2010). Pourquoi coopérer avec un concurrent ? *Revue francaise de gestion*, *204*(5), 155-169.

*Figari, G. (2008). L'évaluation des dispositifs éducatifs. *Mesure et évaluation en éducation*, *31*(3), 77-93. https://doi.org/10.7202/1024966ar

*Garon, R. et Théorêt, M. (2005). Re-connaître les plans à cas unique en sciences de l'éducation. *Mesure et évaluation en éducation*, *28*(1), 1-18. https://doi.org/10.7202/1087725ar

Graber, M. et Haberey-Knuessi, V. (2017). Le bachelor en soins infirmiers : entre professionnalisation et professionnalité en Suisse et en Belgique. *Recherche en soins infirmiers*, *128*(1), 66-78.

*Grangeat, M. (2002). Dix thèses se référant à l'évaluation et récemment soutenues en France : diversité des thèmes et des méthodes, centration sur l'école obligatoire et ses professionnels. *Mesure et évaluation en éducation*, *25*(2-3), 35-40. https://doi.org/10.7202/1088320ar

Gremion, C. (2025). Évancipation : de la mise en dialogue de trois concepts. Dans J.-F. Marcel et C. Gremion (dir.), *Évancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation*. (p. 21-40). Cépaduès-Éditions.

Gremion, C., Burgi, V., Gatti, R. et Le Roy, V. (2021). Identification des indicateurs de professionnalisation dans une formation initiale d'enseignants. Dans C. Gremion et C. de Paor (dir.), *Processus et finalités de la professionnalisation. Comment évaluer la professionnalité émergente ?* (p. 267-291). De Boeck Supérieur.

*Gremion, C. et Coen, P.-F. (2016). De l'influence du contrôle dans les dispositifs d'accompagnement. *e-JIREF*, *1*(3), 11-27.

*Jorro, A. (2004). Réflexivité et auto-évaluation dans les pratiques enseignantes. *Revue Mesure et évaluation en éducation*, *27*(2), 33-47.

Jorro, A. (dir.). (2014). *Dictionnaire des concepts de la professionnalisation* (1è éd.). De Boeck.

Ketele, J.-M. D. et Jorro, A. (2013). *L'engagement professionnel en éducation et formation* (1è éd.). De Boeck supérieur.

*Luisoni, M. et Monnard, I. (2018). Pratiquer l'évaluation sommative au sortir de la formation initiale : quelles tensions, quelles prescriptions ? *e-JIREF*, *4*(1), 67-88. https://doi.org/10.48782/bsm7cq88

*Manier, C. et Abernot, Y. (2016). Contribution à l'évaluation de l'activité professionnelle sur la base d'une CAH : le cas des éducateurs spécialisés. *Mesure et évaluation en éducation*, *39*(2), 39-58. https://doi.org/10.7202/1038241ar

*Maubant, P. (2020). Chapitre 8. Les pratiques évaluatives en contexte de formation par alternance : entre reliance et accompagnement. Dans N. Younès, C. Gremion et E. Sylvestre (dir.), *Evaluations, sources de synergies* (p. 199-231). Les Presses de l'ADMEE.

Maubant, P. et Gremion, C. (2022). Les configurations plurielles de la pédagogie de, par, en (l') alternance. *Phronesis*, *11*(1-2), 1-10. https://doi.org/10.7202/1087554ar

Maubant, P., Roger, L., Caselles-Desjardin, B., Mercier, B. et Gravel, N. (2010). Questions et enjeux dans l'étude compréhensive des processus de professionnalisation dans les métiers de l'intervention humaine. *Les dossiers des sciences de l'éducation*, (24), 109-132. https://doi.org/10.4000/dse.949

*Moukaddam, D. (2024). L'évaluation à visée émancipatrice de la compétence éthique dans la formation enseignante. *e-JIREF*, *10*(3), 5-26. https://doi.org/10.48782/e-jiref-10-3-5

*Neuhaus, O. (2022). L'évaluation croisée entre pairs pour accompagner l'entrée dans le métier des enseignants professionnels. *e-JIREF*, *8*(3), 93-115. https://doi.org/10.48782/w4twg376

Paseka, A., Schratz, M. et Schrittesser, I. (2010). Professionstheoretische Grundlagen und thematische Annäherung. Dans M. Schratz, A. Paseka et I. Schrittesser (dir.), *Pädagogische Professionalität:*

querdenken – umdenken – neu denken: Impulse für next practice im Lehrerberuf (p. 8-45). Facultas Verlags- und Buchhandels AG.

Perrenoud, P. (1993). Formation initiale des maîtres et professionnalisation du métier. *Revue des sciences de l'éducation*, *19*(1), 58-76.

*Perrenoud, P. (1996). Évaluer les réformes scolaires, est-ce bien raisonnable ? *Mesure et évaluation en éducation*, *19*(2), 53-97. https://doi.org/10.7202/1091524ar

Piot, T. (2009). Quels indicateurs pour mesurer le développement professionnel dans les métiers adressés à autrui ? *Questions Vives. Recherches en éducation*, *5*(11), 259-275. https://doi.org/10.4000/questionsvives.622

*Rouiller, J. et Pillonel, M. (2004). Conception des pratiques auto-évaluatives développées par les formateurs en soins infirmiers et en enseignement primaire dans le suivi des stages professionnels. *Revue Mesure et évaluation en éducation*, *27*(2), 49-68.

*Vaudroz, C. et Berger, J.-L. (2018). Validation de la version française de l'Échelle du sentiment de responsabilité des enseignants (Teacher Responsibility Scale). *Mesure et évaluation en éducation*, *41*(3), 87-117. https://doi.org/10.7202/1065166ar

*Vial, M. (1998). Faire participer l'étudiant à l'évaluation de son cursus ? Quelle place pour l'enquête de satisfaction ? *Mesure et évaluation en éducation*, *21*(3), 43-66. https://doi.org/10.7202/1091274ar

*Vial, M. (2006). Les relations entre formation et évaluation : perspectives de recherches. *Mesure et évaluation en éducation*, *29*(1), 81-97. https://doi.org/10.7202/1086969ar

Wittorski, R. (2014). Professionnalisation. Dans A. Jorro (dir.), *Dictionnaire des concepts de la professionnalisation* (1è éd., p. 233-236). De Boeck

Chapitre 15. Réseau thématique international Reconnaissance, valorisation et validation des acquis de l´expérience de l´ADMEE-Europe

Dynamique d'un groupe de recherche et orientations épistémologiques, théoriques et empiriques des investigations

Carmen Cavaco[1] et Pascal Lafont[2]

I. Introduction

La thématique de la reconnaissance et de la validation des acquis de l'expérience a gagné en visibilité sociale à la fin des années 1990, sous l'impulsion d'orientations politiques tant supranationales que nationales. Les politiques, les acteurs, les dispositifs relevant de la reconnaissance, de la valorisation et de la validation des acquis de l'expérience (RVVAE) visent à promouvoir l'obtention d´un diplôme ou d´une partie de diplôme

[1] Instituto de Educação, Universidade de Lisboa, UIDEF.

[2] Université Paris Est Créteil - Val de Marne (UPEC), UFR Sciences de l'Éducation et des Sciences Sociales - SESS/STAPS, Membre du Laboratoire LIRTES EA 7313.

en faisant valoir les acquis de l´expérience. Ce contexte a favorisé le développement de pratiques et la réalisation de recherches sur le sujet. Les pratiques de reconnaissance des acquis de l'expérience reposent principalement sur deux présupposés : d'une part, que l'expérience possède un potentiel formatif et qu'elle donne lieu à des apprentissages ; d'autre part, la conviction qu'il est non seulement possible, mais également nécessaire et souhaitable, de rendre ces acquis socialement visibles. Dans ce contexte, la création d'un réseau thématique international portant sur la Reconnaissance, valorisation et validation des acquis de l'expérience au sein des activités de l'ADMEE-Europe, s'est imposée se donnant pour objectif de travailler les dimensions épistémologiques, théoriques et éthiques soulevées par l'évaluation inhérente aux dispositifs de RVVAE.

L'un des postulats partagés par l'ensemble des membres de ce Réseau conduit à considérer que les dispositifs de RVVAE comportent une dimension d'évaluation qui ne peut être considérée à l'égale de celle produite dans les processus d'évaluation traditionnellement observés en milieu scolaire. En effet, l'évaluation réalisée dans les pratiques relatives à la RVVAE présente des spécificités qui parfois sont négligées, voire ignorées par les acteurs impliqués en raison du risque de perte de leurs prérogatives ou de la remise en question de leurs pratiques habituelles. Il convient donc de rappeler que, si les dispositifs de RVVAE peuvent s'appuyer sur des principes de justice sociale dans la mesure où ils constituent une voie alternative d'accès à la formation et à la certification pour des personnes n'ayant pas eu d'autres opportunités, ils peuvent également être confrontés à des logiques et pratiques d'évaluation qui remettent en cause et subvertissent ces principes.

C'est pourquoi ce chapitre a pour vocation, d'une part, de montrer comment des questions vives et leur traitement ont permis l'émergence et le renforcement d'une dynamique d'un groupe de recherche en développant une activité scientifique constante depuis sa création et, d'autre part, d'identifier les productions collectives de ce Réseau ainsi que certaines dimensions épistémologiques, théoriques et éthiques structurantes de l'évaluation des acquis de l´expérience

II. Construction et dynamique d'un groupe thématique de recherche

Si l'expérience est la vie même, comme l'écrit Dewey (1968), elle est donc tout ce qu'une personne a fait, ce qu'elle a été capable de faire, ce qu'elle n'a pas fait, pas pu, pas su ou pas voulu faire. Elle est par conséquent aussi ce potentiel de développement pour les expériences à venir. Bien que simplificateur, ce constat suffit à rappeler la complexité que les chercheurs et praticiens de ce Réseau ont à affronter quand ils travaillent avec l'expérience. Il s'agit donc de faire référence au paradigme de la complexité pour analyser et problématiser les activités et pratiques évaluatives relatives à la RVVAE car « pour penser la complexité il faut modifier nos schémas mentaux, changer de paradigme en passant de la pensée simplifiante à la pensée complexe » (Lazzari Dodeler et al., 2023, p.2).

II.a. Constitution évolutive du groupe de recherche

Le réseau thématique Reconnaissance, valorisation et validation des acquis de l'expérience (RVVAE), créé en 2002 par Gérard Figari, est un groupe de recherche qui s'est donné pour finalité de partager, de diffuser et de faire progresser d'une part, la recherche sur les procédures d'évaluation mobilisées dans les dispositifs de reconnaissance et de validation des acquis de l'expérience ; et d'autre part, les connaissances sur les dispositifs de reconnaissance, de valorisation et de validation de l'expérience, notamment rapportées aux enjeux liés à la prise en compte de l'expérience. D'ailleurs, cette prise en compte peut être considérée comme un analyseur des transformations sociales et sociétales actuelles ; au sens où « ainsi, dénommées, un ensemble d'évolutions caractérisent tendanciellement la société contemporaine, se répercutent sur les activités humaines, notamment au travail et en formation, et redessinent les relations entre éducation, formation et travail » (Champy-Remoussenard, 2015, p.16). Aussi, ces évolutions sont-elles à considérer au prisme de la (re)mise en question de l'évaluation elle-même, ainsi que des activités et des outils d'évaluation, face aux défis posés par la RVVAE.

Ce réseau thématique RVVAE de l'ADMEE-Europe a été coordonné au fil du temps par : Gérard Figari, Patrick Mayen, Marie-Christine Presse, Carmen Cavaco et Pascal Lafont. Le Réseau dispose d'une structure très souple, tant dans ses modes d'organisation et de fonctionnement que dans sa composition, à la fois variable et ouverte, dans la mesure où tous les chercheurs et praticiens intéressés à collaborer peuvent l´intégrer. Le renouvellement de ses membres par un flux d'entrées et de sorties

demeure une préoccupation permanente, ce qui constitue un principe grâce auquel de nouvelles recherches sont valorisées, de nouveaux collègues sont reconnus. Ce réseau regroupe des membres issus de plusieurs pays, dans lesquels existent des politiques et des pratiques de RVVAE. L'évolution du périmètre et une diversité dans la constitution du réseau ont permis le développement d'activités scientifiques auxquelles ont contribué des membres[3] venant, notamment, de la Belgique, du Brésil, du Canada, de la France, du Portugal, du Royaume-Uni, du Liban et de la Suisse.

La création et la dynamique du Réseau RVVAE reflètent certaines des grandes qualités de Gérard Figari : sa capacité à initier et à fédérer des projets collectifs, à valoriser la dimension internationale ainsi que la diversité et la complémentarité des points de vue, à reconnaître et à encourager le travail de jeunes chercheurs, et à conjuguer rigueur, convivialité, sens du dialogue et humanisme. Peu après la création de ce réseau, en 2002, un colloque de l'ADMEE-Europe, en partenariat avec l'AFIRSE Portugal, a été organisé en 2004 portant sur le thème de l'évaluation des compétences : entre reconnaissance et validation des acquis de l'expérience. La création du Réseau, rapidement suivie par la tenue de ce colloque, alors que Gérard Figari présidait l'ADMEE-Europe, illustre la volonté qui fut la sienne d'impulser cette thématique, laquelle, selon lui, permettrait d'élargir le champ de l'évaluation et d'en dépasser les conceptions traditionnelles.

II.b. Une activité scientifique régulière et structurante

[3] Alain Savoyant, Anne Bera, Anne Grzyb, Anne Jorro, Anne-Gaëlle Dorval, Arnaud Salmon, Bernardine Rivoire, Bohrène Chakroun, Brigitte Pagnani, Carmen Cavaco, Cathal de Paor, Catherine Mayeux, Claire Tourmen, Edmée Ollagnier, Emmanuel Triby, Françoise De Viron, Geneviève Lameul, Geneviève Tschopp, Gérard Figari, Gilles Leclerc, Helen Murphy, Hervé Breton, Isabelle Houot, Jacqueline Beckers, Jean Matter, Jean-François Métral, Jean-Marc Braibant, Jean-Marie De Ketele, Jérôme Eneau, José Montéagudo, Kristine Balslev, Louise Belair, Marcel Pariat, Marie-Christine Presse, Marie-Christine Vermelle, Marie-Claire Dauvisis, Matthis Behrens, Mireille Pouget, Muriel Joseph-Théodore, Nina Asloum, Noëlle Lefebvre, Palmira Alves, Pascal Lafont, Pat Davies, Patrick Mayen, Patrick Rywalski, Philippe Mougel, Pierre Alain Besaçon, Pierre Valois, Sandrine Cortessis, Saskia Weber Guisan, Simone Valdete dos Santos, Sonia Fradette, Suzanne De Jonckeere, Yves Obré, ainsi que d'autres collègues.

Au cours des vingt-trois années d'existence de ce réseau thématique, les échanges et discussions se sont principalement déroulés dans le cadre des symposiums intégrés aux colloques de l'ADMEE-Europe et, occasionnellement, lors de symposiums organisés dans le cadre d'autres événements scientifiques, tels que la Biennale de l'Éducation et de la Formation, l'AFIRSE Portugal et le groupe pour l'évaluation des pratiques professionnelles (gEvaPP). A l'issue de ces moments de partage et de discussion collective lors de symposiums, les membres du Réseau saisissaient l'occasion de tenir une réunion interne en vue de réaliser une synthèse des réflexions destinée à l'ensemble des membres du réseau et de planifier ses activités.

Entre 2002 et 2024, le Groupe a organisé 14 symposiums, une table ronde et un séminaire lors des Colloques de l'ADMEE-Europe et d'autres entités partenaires, afin de discuter et de diffuser les connaissances scientifiques produites par la recherche de ces chercheurs sur les spécificités, la complexité et les enjeux de l'évaluation à partir des acquis expérientiels et des compétences, mais aussi sur l'épistémologie de ce type d'évaluation, les méthodologies et des outils, les référentiels, la formation, la posture et les pratiques évaluatives des acteurs impliqués (accompagnateurs et jurys), ainsi que les dimensions éthiques et déontologiques soulevées par ces politiques et pratiques.

La dynamique collective a permis la création et la consolidation d'un réseau de chercheurs et de praticiens dans le domaine des sciences de l'éducation, autour de thèmes de recherche communs (l'éducation des adultes, la reconnaissance, la valorisation et la validation des acquis de l'expérience, ainsi que l'évaluation) en adoptant une perspective interdisciplinaire. Outre cette dynamique et les productions mises en évidence, ce collectif a commencé à travailler de manière collaborative à d'autres occasions et dans d'autres projets, autour de la diffusion des résultats de recherche lors de symposiums (REF et Biennale de l'Éducation et de la Formation), ainsi que dans des projets collectifs de publication et de recherche. Ainsi, la dynamique présentée dans ce texte ne constitue qu'une partie du travail accompli, car, au fil de ce processus et au cours de cette période, des partenariats entre chercheurs ont émergé et se sont consolidés, donnant lieu à divers projets collaboratifs.

III. Productions scientifiques et éléments d'analyse

III.a. Productions scientifiques

Quels que soient les ancrages théoriques et disciplinaires, dans leur ensemble, les productions de ce réseau RVVAE ont toujours poursuivi le but de faire coexister une dominante théorique et une dominante empirique de façon à nourrir le principe d'un rapport dialectique entre les deux dimensions.

Au début du travail du Réseau, Matthis Behrens, Jean-Marc Braibant et Marie-Claire Dauvisis (2004) ont réalisé une synthèse des travaux menés dans divers pays, afin de mettre en exergue l'état des recherches centrées sur la reconnaissance et la validation des acquis de l'expérience (RVAE), et de fournir une base pour des théorisations, des questionnements et des approfondissements concernant les pratiques et les recherches en évaluation. Ce travail a permis d'élaborer une analyse qui a déclenché la poursuite des activités du Réseau, en mettant en évidence les thèmes transversaux des recherches : 1) Les expériences en matière de RVAE varient fortement d'un pays à l'autre. Le cadrage législatif détermine l'architecture des dispositifs et la nature des pratiques ; 2) Les procédures de RVAE sont souvent complexes ; 3) L'intérêt pour la RVAE est toujours double : développement personnel et accès à la certification ; 4) Un problème central de la VAE est la qualité docimologique des démarches proposées ; 5) Il n'est pas certain que l'on puisse décrire la complexité des codes et des savoirs du monde académique au moyen d'une approche par compétences ; 6) La RVAE est tributaire des aléas du marché du travail, des attentes, intérêts et enjeux de tous les acteurs impliqués.

En 2010, Marie-Christine Presse et Gérard Figari ont dirigé le numéro 4 de la revue *Transformations*. Recherches en éducation des adultes, consacré à la valorisation des expériences personnelles et professionnelles des adultes en Belgique, en France, au Portugal et en Suisse. Les différentes contributions, rédigées par des chercheurs du Réseau, explorent les problématiques et enjeux majeurs qui se jouent autour de la valorisation des expériences personnelles et professionnelles des adultes, mettent en évidence des tensions, contradictions et difficultés, ayant de profondes répercussions sur l'organisation du dispositif, sur les méthodologies, sur les instruments, ainsi que sur les fonctions et attitudes des acteurs impliqués. Les articles se concentrent autour de deux questions principales : quelles sont les pratiques mises en œuvre permettant aux

candidats engagés dans ces processus de valoriser leurs expériences et de les faire reconnaître ? Quelles sont les modalités de modélisation de l'expérience, de référentialisation, de définition des critères et indicateurs de compétences mobilisées dans ces procédures ? Les recherches mettent en évidence la diversité et la complexité de ces pratiques et soulignent l'importance de réfléchir au statut épistémologique de l'évaluation et d'y intégrer une approche interdisciplinaire et située. Ces travaux permettent l'émergence des problématiques de recherche et de modélisations : 1) les pôles de tension dérivés de la coexistence de différentes logiques (personnelle, économique, politique, sociale et des experts) ; 2) l'analyse de la spécificité du processus à partir des notions de (re)élaboration de l'expérience, de transformation de l'expérience et de double processus d'élaboration, ainsi que les notions de distance et de distanciation par rapport aux acquis de l'expérience opérés par les différents acteurs tout au long du processus ; 3) les compétences en jeu, comme les compétences de raisonnement, les compétences d'explicitation de l'action, les compétences de lecture et d'écriture, et les compétences cognitives, relationnelles et maîtrise de la langue ; 4) le rôle et le statut des référentiels, entre le normatif et le prescriptif, et une approche plus critique et flexible envisageant leur adaptation afin d'assurer leur sens et leur utilité pour les acteurs impliqués. Ces recherches mettent en évidence les particularités évaluatives des procédures de reconnaissance et de validation, tout en montrant que la prise en compte de l'expérience constitue une nouvelle manière d'évaluer, en rupture avec les pratiques héritées du système scolaire.

En 2020, Pascal Lafont et Carmen Cavaco ont publié un numéro hors-série dans la revue *Chemins de formation*, consacré aux « défis de l'évaluation dans les pratiques de Reconnaissance et de Validation des Acquis de l'Expérience » dans lequel il s'est agi d'appréhender l'origine de ces défis et leur traduction dans la détermination des actions produites et induites par les acteurs concernés, à partir de recherches scientifiques développées dans le champ des sciences de l´éducation dans divers pays d´Europe (France, Portugal, Suisse, Belgique, Irlande). Prenant appui sur des données factuelles et théoriques, les contributeurs ont mis en débat les tensions et les paradoxes issus des positionnements et du sens de l'action sous-tendus par des pratiques d'évaluation, en vue, à la fois, d'analyser les défis les plus significatifs au regard de l'usage du portfolio, des référentiels, des preuves et de l´observation de l´activité, et de rendre compte, entre savoirs et compétences, de ce qui est finalement évalué au travers des pratiques de RVVAE. Les contributions portent des résultats démontrant comment chacun des acteurs de l'évaluation des acquis de

l'expérience se prépare à participer et mettre en jeu ses conceptions, et plus encore sa position au sein de la situation évaluative. Si les tensions peuvent marquer tant de préoccupations communes, il convient d'observer que « ce qui fait valeur relève de principes d'action ou de métier donc d'une valeur professionnelle, mais également de dynamiques d'organisation voire de systèmes et rejoindre ainsi la valeur économique. La démarche de VAE peut donc simultanément constituer l'évaluation d'un savoir nouveau et la reconnaissance d'un potentiel de développement à venir » (Triby, 2020, p.63).

Au cours de la même année (2020), Pascal Lafont, Patrick Rywalski et Carmen Cavaco ont produit un chapitre d'ouvrage, intitulé, « Évaluation sommative versus évaluation formative dans les pratiques d'évaluation des acquis de l'expérience » dans le cadre du lancement des Presses de l'ADMÉE. Cette production a rendu compte des travaux réalisés dans le contexte d'un symposium qui s'est donné pour objectif d'interroger la complémentarité ou l'opposition entre les modalités, les processus, les postures, les dispositifs et les formes d'évaluation ainsi que leur mise en œuvre dans les pratiques d'éducation et de formation professionnelle des jeunes et des adultes. L'évaluation des acquis non formels et informels dépend très étroitement de l'usage des outils de mesure (référentiels) et du poids de la preuve (ou évidences) d'autant plus que « l'intérêt des outils est celui de l'unification des modalités d'évaluation, dans le principe majeur qu'est l'égalité de traitement des candidats sur le territoire national » (Lainé, 2018, p. 72). Des obstacles culturels œuvrent encore contre la reconnaissance des savoirs d'action et conduisent à légitimer le recours à l'évaluation certificative au détriment de l'évaluation formative. Ainsi, le réseau a-t-il travaillé les problématiques suivantes : Quels sont alors les défis relatifs au développement de l´évaluation formative dans les pratiques d'éducation et de formation des jeunes et des adultes ? Quels outils portent l´évaluation formative en faveur de la reconnaissance des acquis d'apprentissage ? L'évaluation associée aux fonctions de normalisation et de régulation constitue-t-elle un frein à l´usage de l´évaluation formative (Thélot, 2008) ? Ces questions et leurs réponses peuvent-elles être orientées par des évolutions contextuelles ?

En 2023, Pascal Lafont et Sandrine Cortessis ont dirigé un numéro de la revue *L'orientation scolaire et professionnelle*, intitulé « Mise en mots et évaluation de l'expérience : entre institutionnalisation et émancipation », qui défend l'idée que les écrits expérientiels permettent à leur auteur de développer des compétences méta-réflexives et d'autoanalyse, favorisant l'articulation entre les « savoirs enseignés » et « les savoirs d'expérience »

(Barbier et Galatanu, 1996) ainsi que de prendre conscience ou de construire une nouvelle identité professionnelle et/ou sociale. La discussion est soutenue par l'analyse de récits d'expériences tant scolaires qu'universitaires, professionnelles et sociales, individuelles et collectives à l'aune de la complexité et de la pluralité des évaluations concernant la reconnaissance et la validation des acquis de l'expérience dans des contextes divers. La réflexion collective a conduit à traiter la question suivante : à quelles conditions les récits d'expérience peuvent-ils ouvrir des perspectives d'émancipation malgré le cadre institutionnel qui en prescrit et en circonscrit l'usage ? Les résultats des différentes contributions montrent que la mise en récit de son expérience conduit à apprendre sur soi, et que la transmission des analyses appliquées à son expérience vers autrui à partir de son propre récit d'expérience constitue une condition d'acquisition de compétences valorisées et valorisables en matière de RVVAE. Par la confrontation de son expérience transformée à celle des autres, de sa propre expérience antérieure au processus d'analyse induit par la RVVAE, une prise de conscience, étape déterminante dans tout processus d'apprentissage et de construction identitaire, s'opère à travers un travail de triple transaction biographique-relationnelle-intégrative (Perez-Roux, 2019).

III.b. Questionnements épistémologiques et orientations des recherches

Travailler sur les modalités de RVVAE revient à interroger leur nature et leur légitimité. Reconnaître la valeur épistémologique de l´expérience, en tant que possibilité de développer des savoirs sous-tend le travail et les questionnements épistémologiques du Réseau RVVAE. Cela implique de rompre avec une vision strictement scolaire ou disciplinaire et de reconnaître des formes de savoir non codifiées. Ainsi, la question se pose de problématiser ce qui est considéré comme des savoirs valides : savoirs académiques, expérientiels, pratiques, tacites, implicites ou culturels. La nature et la spécificité des savoirs de l´expérience relève d'une dimension incontournable à prendre en compte pour saisir le sens et comprendre les résultats de la recherche en matière de RVVAE. Ces savoirs sont contextualisés, incorporés, globaux et difficilement énonçables (Mayen et Mayeux, 2003). Autrement dit, comment évaluer les acquis de l´expérience à partir de leur nature et de leur spécificité ?

La problématique relative à l'explicitation des acquis d'expérience, nécessaire à leur reconnaissance et validation, a fait l'objet de présentations et de débats. Bien que ce processus soit complexe en raison de la nature incorporée, composite et difficilement énonçable des acquis issus de

l'expérience, il se confronte aux référentiels, aux méthodologies et aux outils d'évaluation, qui peuvent orienter le processus vers une évaluation sélective, régulatrice, normative et certificative, ou, au contraire, vers une évaluation favorisant la « réflexivité » (Jorro et Droyer, 2019), autrement dit une « évaluation pour apprendre » (Mottier Lopez et al., 2021).

La dimension herméneutique et interprétative a fait l'objet de nombreuses contributions, en ce sens qu'elle structure le processus d'explicitation, de socialisation et d'évaluation des acquis de l'expérience. L'adulte candidat explicite ses acquis à partir d'un processus de mise en récit et de (re)élaboration de l´expérience. Parallèlement, l'évaluateur assume un rôle d'interprète, portant un regard sur le parcours du candidat, ce qui requiert des compétences pour comprendre les contextes, les significations et les transferts de savoirs. Les recherches problématisent les interdépendances entre interprétation et mise en récit de l'expérience, au prisme, par exemple, des portfolios et des interactions entre les acteurs impliqués (adultes, accompagnateurs et évaluateurs).

Le rôle social et politique de l'évaluation constitue également un axe de travail. L'évaluation n'est pas neutre : elle confère une valeur sociale aux savoirs et peut (ou non) favoriser la reconnaissance, la formation, la justice sociale et la mobilité sociale et professionnelle. C'est pourquoi d'autres questionnements ont émergé pour un traitement qui a conditionné la dynamique du groupe de recherche à travers le déploiement de discussions scientifiques, riches et stimulantes. La RVVAE soulève des questions épistémologiques, méthodologiques, théoriques, éthiques et déontologiques. Leur analyse prend en compte la complexité des situations, aussi plurielles que celles que présentent les échelles locales, régionales, nationales, européennes, voire internationales, des discours et des pratiques en matière de RVVAE.

Des questionnements fondamentaux ont constitué un fil conducteur des réflexions : définition et nature de l'expérience (Qu'est-ce que l'expérience et quelle est la nature des acquis de l'expérience et des compétences ?) ; rapports entre savoirs expérientiels et savoirs académiques/scientifiques (Comment mettre en rapport les savoirs issus de l'expérience avec les savoirs académiques/scientifiques dans les pratiques d'évaluation ?) ; reconnaissance et évaluation des acquis de l'expérience (De quelle manière l'évaluation est-elle présente dans la reconnaissance, la valorisation et la validation des acquis de l'expérience ? Comment mettre en œuvre des pratiques évaluatives capables de reconnaître, valoriser et valider les acquis de l'expérience tout en prenant en compte leur complexité et singularité ? Qu'évaluons-nous : les acquis de l'expérience, les

compétences, la capacité réflexive ou la mise en mots ?) ; rôle et limites des référentiels et des outils (Quels rôles jouent les référentiels et les outils : médiateurs favorisant la reconnaissance, la valorisation et la validation des acquis de l'expérience, ou cadres réducteurs qui engendrent des défis et des tensions ?) ; catégories de professionnels impliqués – formation, fonctions et postures (Quels rôles et défis pour les différents acteurs impliqués - accompagnateurs, formateurs, jurys ? Quels repères, instruments et principes éthiques guident ces pratiques ? Quelles sont les spécificités de leur formation, de leur agir professionnel et de leur déontologie ?). Le Réseaux a progressé dans la construction de connaissances sur ce sujet, mais il reste important d'investir dans l'approfondissement de ces questions vives de recherche.

A partir de ces questions, le Réseau a problématisé la (re)mise en question de l'évaluation elle-même, des activités et outils de l'évaluation face au défi posé par la reconnaissance, la valorisation et la validation des acquis de l'expérience. L'activité évaluative inhérente aux pratiques de reconnaissance, de valorisation et de validation des acquis de l'expérience implique des jugements de valeur de la part des acteurs concernés (adultes, accompagnateurs et jurys), consistant à comparer des repères expérientiels (acquis de l'expérience) à des référents (référentiels) (Figari et Tourmen, 2006). L'évaluation a été réinterrogée à partir de l'analyse de l'émergence, de la construction et de la réalisation de ces activités, selon les cadres épistémologique, théorique, méthodologique et la dimension éthique. Ces questionnements ont contribué à problématiser l'émergence d'une nouvelle forme d'évaluation dans les processus de reconnaissance, de valorisation et de validation des acquis de l'expérience, à comprendre les spécificités de cette « évaluation située » (Figari, 2010), ainsi qu'à favoriser la production de nouveaux savoirs sur l'expérience et sur les compétences. Dans cette activité évaluative, un processus de (re)construction des référents et des référés serait à l´œuvre pour permettre de formuler des interprétations (Figari et Tourmen, 2006).

III.c. Dimensions éthiques

La spécificité, la complexité et les tensions qui caractérisent le processus de reconnaissance, de valorisation et de validation des connaissances acquises par l'expérience soulèvent de nombreuses questions d'ordre éthique, lesquelles ont été abordées par le Réseau dans ses discussions et publications. Les pratiques évaluatives inhérentes à ce type de processus posent des questions quant à l'instrumentalisation de la finalité éducative au service d'objectifs économiques et parfois sociaux, tout en occultant la question de la justice sociale, de l´épanouissement du sujet et de

l´émancipation. La question du rôle et de la posture des acteurs impliqués dans les processus de RVVAE reste ouverte. Les candidats, accompagnateurs, conseillers VAE et jurys se trouvent au cœur de ce processus, souvent partagés entre différentes tensions. Ils sont, entre autres fonctions, responsables du déroulement complet du processus, depuis les orientations initiales, en passant par l'accueil des candidats, la formalisation des contrats, l'émergence des compétences, la rédaction du dossier et, enfin, sa validation par un jury.

Évaluer les acquis de l'expérience est un processus très exigeant et difficile pour tous les acteurs impliqués, compte tenu de la nature située, composite et implicite des savoirs issus de l'expérience. Cela entraîne des répercussions sur le travail accompli par les adultes, les accompagnateurs et les jurys. Lorsque l'adulte parvient à obtenir une validation totale, les conséquences sur son estime de soi et sur son identité sont très positives. En revanche, lorsque sa démarche aboutit à une validation partielle ou à un refus de validation de la part du jury, les effets peuvent être contre-productifs pour l'adulte, qui a alors tendance à se sentir responsable de l'échec, considérant que son expérience de vie et ses savoirs n'ont aucune valeur. Face à la spécificité et à la complexité, l'évaluation située qui se déroule dans cette démarche exige de chaque acteur une posture éthique, afin d'assurer un processus fondé sur le respect d'autrui, l'empathie raisonnée, la reconnaissance, la dignité, la transparence et la rigueur. Cette posture exige une conscience permettant d'observer comment les pensées et les réflexions peuvent être affectées par les états corporels (Linard, 2003) au sens où ils peuvent rendre compte d'émotions et de perceptions, restituer une représentation de soi et d'un devoir d'agir.

IV. Conclusion

Le Réseau thématique international Reconnaissance, valorisation et validation des acquis de l'expérience de l'ADMEE-Europe a contribué à l'analyse et à la problématisation de la spécificité, de la complexité et des tensions qui caractérisent l'évaluation située et réalisée dans le cadre de la reconnaissance, de la valorisation et de la validation des acquis de l'expérience, dans leurs dimensions épistémologiques, théoriques, méthodologiques et éthiques. Les recherches menées ont mis en évidence que, bien que le processus de reconnaissance, de valorisation et de validation ne se limite pas à l'évaluation, un important processus évaluatif y est sous-jacent. Toutefois, cela n'est pas toujours reconnu et assumé, notamment par les accompagnateurs et les membres du jury. La formation

de ces professionnels, notamment dans le domaine de l'évaluation, constitue une stratégie essentielle pour leur professionnalisation. Elle permet de leur assurer l'acquisition des connaissances nécessaires et appropriées à l'exercice de leurs fonctions, d'élargir leur « conception du processus évaluatif » (De Ketele, 2010, p. 25) et de comprendre, maintenir et adapter leur posture professionnelle ainsi que les dimensions éthiques aux spécificités et à la complexité du processus. Malgré les avancées réalisées dans ce domaine et des pratiques d'évaluation analysées à tous les niveaux de certification (Lafont et Pariat, 2024), le Réseau a encore un rôle important à jouer dans la promotion de la recherche, le partage et la diffusion des connaissances relatives à l'évaluation dans le processus de RVVAE. Cela concerne notamment la nature de l'expérience et des savoirs expérientiels, les modalités de reconnaissance et de validation, les référentiels, l'évaluation des compétences, les méthodologies et les instruments, les finalités, l'organisation et le fonctionnement du processus, l'accompagnement et la formation des acteurs impliqués, ainsi que les tensions, les défis et les dimensions éthiques. Ces éléments permettront d'avancer dans la consolidation de ce que Gérard Figari (2010) désignait comme les « plusieurs facettes d'un paradigme en émergence : l'évaluation de l'expérience » (p. 136). Toutefois, face à la sollicitation d'organismes employeurs, à des logiques d'employabilité et de qualification invitant les candidats potentiels à utiliser de nouveaux outils issus de l'intelligence artificielle, de nouvelles questions émergent tout comme de nouveaux défis. Il s'agit donc de repenser les formes d'évaluation afin de garantir l'authenticité de la narration de l'expérience, la profondeur de réflexion et la pertinence de l'accompagnement face aux outils numériques émergents.

V. Bibliographie

Barbier, J.-M. et Galatanu, O. (2004). Les savoirs d'action : une mise en mot des compétences ? L'Harmattan.

Behrens, M., Braibant, J.-M., et Dauvisis, M.-C. (2004). Réseau de reconnaissance et de validation des acquis de l'expérience (RVAE) de l'ADMEE : Synthèse des travaux réalisés dans trois axes et questionnements propres à faire avancer le débat en vue du colloque de Lisbonne (Document polycopié). Communication présentée au XVII[e] colloque ADMEE-Europe, 18–20 novembre 2004, « L'évaluation des compétences entre reconnaissance et validation des acquis de l'expérience ».

Lisbonne : Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa.

Champy-Remoussenard, P. (2016). Les transformations des relations entre travail, éducation et formation dans l'organisation sociale contemporaine : questions posées par trois dispositifs analyseurs. Revue française de pédagogie, (190), 15-28.

De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. Revue française de linguistique appliquée, XV(1), 25-37.

Dewey, J. (1968). Expérience et éducation. Armand Colin.

Figari, G. (2010). Conclusion. Transformations. Recherches en éducation des adultes, CUEEP, (4), 131-137.

Figari, G. et Tourmen, C. (2006). La référentialisation : une façon de modéliser l´évaluation de programme, entre théorie et pratique. Vers une comparaison des approches au Québec et en France. Mesure et Évaluation en Éducation, 29(3), 5-25.

Jorro, A. et Droyer, N. (2019). L'évaluation, levier pour l'enseignement et la formation. De Boeck Supérieur.

Lafont, P. et Cavaco, C. (dir.) (2020). Les défis de l'évaluation dans les pratiques de Reconnaissance et de Validation des Acquis de l'Expérience (RVAE). Chemins de Formation. L'Harmattan.

Lafont, P. et Cortessis, S. (dir.) (2023). L'orientation scolaire et professionnelle [En ligne], 52/3, mis en ligne le 01 septembre 2025. https://journals.openedition.org/osp/17723

Lafont, P. et Pariat, M. (dir.) (2024). Doctorat et Validation des acquis de l'Expérience : De la certification académique au titre professionnel. Editions Pétra.

Lafont, P., Rywalski, P. et Cavaco, C. (2020). Évaluation sommative versus évaluation formative dans les pratiques d'évaluation des acquis de l'expérience. Dans N. Younès, C. Gremion et E. Sylvestre (dir.), Évaluations, sources de synergies. Les Presses de l'ADMEE.

Lainé, A. (2018). Evaluer l'expérience en VAE – Le dialogue des gens de métier. Editions Erès.

Lazzari Dodeler**, N,** Albert, M.-N. **et** Tremblay, D.-G. (2023). Simplicité et complexité des crises à la lumière du paradigme de la complexité d'Edgar Morin. *Revue Interventions économiques*, (69), 1-9.

Linard, M. (2023). Autoformation, éthique et technologies : enjeux et paradoxes de l'autonomie. Dans B. Albero (dir.), Autoformation et enseignement supérieur (p. 241-263). Hermès-Lavoisier.

Mayen, P. et Mayeux, C. (2003). Expérience et formation. Savoirs, (1), 13-53.

Mottier Lopez, L., Girardet, C. et Naji, T. (2021). L'évaluation continue pour apprendre : enjeux de la pluralité des feedbacks entre pairs dans un cours universitaire. Mesure et évaluation en éducation, 44(2), 1-128.

Perez-Roux, M. T. (2019). Devenir enseignant en seconde carrière : des reconversions désirées à l'épreuve du réel. Recherche et Formation, (90), 27-41.

Presse, M.-C. et Figari, G. (dir.) (2010). La valorisation des expériences personnelles et professionnelles, Transformations. Recherches en éducation des adultes, (4). CUEEP. https://transformations.univ-lille.fr/index.php/TF/issue/view/6

Thélot, C. (2008). L'évaluation dans le système éducatif : pour quoi faire ? A quelles conditions l'évaluation est-elle utile ? Dans G. Baillat, J.-M. De Ketele, L. Paquay et C. Thélot (dir.), Evaluer pour former (p. 17-27). De Boeck Université.

Triby, E. (2020). L'évaluation en VAE aux limites de l'expérience. Chemins de formation, Hors-série, 53-67.

Chapitre 16. Évaluer en situation de travail, un renouvellement des questions d'évaluation des compétences

Patrick Mayen

I. Introduction

Les questions d'évaluation des compétences en situation de travail ont été posées depuis longtemps dans le cadre des formations dites « duales » ou encore « par alternance ». On les a vues s'infléchir, notamment à l'occasion de la mise en place de la validation des acquis de l'expérience, avec le défi posé par l'évaluation-validation de compétences à partir des écrits ou des propos tenus par les candidats. Ces modalités évaluatives « intermédiaires » se sont aussi développées dans les formations professionnelles, avec une plus grande place accordée à ce qu'on peut appeler l'analyse de situations vécues ou le retour d'expérience après les séquences en milieu professionnel (Tardif, Borges, Malo, 2012). Leur intérêt repose sans doute sur le constat de la grande difficulté d'évaluer des compétences en situation de travail qui ne se limite pas aux compétences pour *la* situation de travail spécifique où se fait l'évaluation. En effet, comment parvenir à évaluer les capacités à généraliser - au moins à un certain niveau - à ajuster l'action en fonction des variabilités des situations de la même classe qui, par essence, ne se retrouvent pas complètement dans la situation spécifique ?

Ces questions viennent aussi de trouver un renouveau avec la mise en place, en France, des Actions de Formation en Situations de Travail (AFEST) (Education Permanente, 2021) L'objectif du texte qui suit vise à montrer que cette nouvelle forme d'action de formation propose de nouveaux défis à l'ingénierie de formation, mais aussi à l'évaluation. L'évaluation peut y être considérée comme un ensemble de moyens structurants pour raisonner et concevoir la formation en situation de travail. Au-delà des spécificités des AFEST, on peut aussi penser qu'elles sont l'occasion de revoir autrement les formations duales ou par alternance ainsi que les questions d'évaluation qui leur sont propres.

II. Évaluer en situation de travail, un renouvellement des questions d'évaluation des compétences

L'idée initiale est de reconnaître le rôle formateur du travail. Une AFEST se distingue de l'apprentissage sur le tas. On peut identifier quelques caractéristiques de celui-ci : il est soumis aux aléas du travail et des tâches confiées à une personne, peut se dérouler avec ou sans aide d'un autre, n'est soumis ni à un contrôle ni à une évaluation. En reprenant une distinction mise en lumière par John Dewey (2012), l'expérience en situation d'apprentissage sur le tas peut aller d'une expérience « eue » à une expérience « faite ». Selon les caractéristiques du travail lui-même dans la situation vécue, selon le rôle joué par l'entourage, selon la conscience et l'intention d'apprendre, selon les compétences préalables de celui qui vit la situation, l'apprentissage et le développement des compétences peuvent être très variables et engendrer des résultats très différents.

En rupture avec l'apprentissage sur le tas, pour être reconnue comme une « action » de formation au sens juridique du terme (dans le système fortement normé de la formation professionnelle en France), une AFEST doit fixer des objectifs explicites, se dérouler avec un « tuteur » ou formateur désigné, donner lieu à une évaluation. En outre, afin que la formation ne soit pas seulement adaptative à une situation spécifique ou à un poste de travail donné, un tel dispositif prévoit deux temps distincts : une mise en situation et une séquence d'analyse réflexive dont le but est de prendre de la distance d'avec la situation. On pourrait reprendre les

concepts de Jean Piaget (1992), pour rendre compte de l'intention de donner comme finalité non seulement « la réussite », mais aussi « la compréhension » des effets de l'action, réussite ou échec. Il s'agit donc de viser non plus la construction de l'exécution de l'action, mais aussi une certaine conceptualisation de celle-ci. Le but est d'étendre le champ de ce qui est acquis à une classe plus large de situations, à construire des compétences pour faire face à d'autres situations de la même famille, à leurs variations et à un degré plus élevé de complexité. Extension à d'autres situations, généralisation, conscience et compréhension des tenants et aboutissants de l'action, capacités de diagnostic de la situation et d'ajustement de l'action en fonction des variations de celle-ci composent ce qu'on peut déjà considérer comme les principaux critères d'une AFEST effective. Ajoutons enfin que, dans le cas des AFEST, il est recommandé, lorsque c'est possible, d'adosser la formation à une unité de certification reconnue.

Une série de difficultés sont rapidement apparues tout au long du processus de conception et de conduite d'une AFEST : définition des compétences à construire et de la situation à maîtriser, évaluation du potentiel formatif de la situation de travail elle-même et des conditions «formatives » à mettre en place, notamment lorsqu'un aménagement des conditions du travail est nécessaire pour rendre la situation réellement formative (cela peut aller jusqu'à sortir de l'entreprise pour trouver les conditions dans une autre entreprise ou dans un atelier d'un centre de formation), mise en place et conduite des moments de réflexivité, évaluation.

On a pu aussi observer des difficultés à traiter, voire tout simplement « penser », les enjeux évaluatifs. Notons que cela n'est pas étonnant puisqu'il en va de même dans de très nombreux cas de la formation ou de l'enseignement professionnel : définir les objectifs, en termes de compétences attendues, évaluer le potentiel formatif des situations de travail à maîtriser, définir les critères et modalités d'évaluation relèvent des difficultés qui apparaissent rapidement au cours de la conception d'une AFEST. Mais les enjeux d'une évaluation-positionnement des personnes qui vont s'engager dans celle-ci, d'une évaluation pendant la formation, formative, mais aussi régulatrice, les précautions pour mettre en place une évaluation dans le cadre particulier d'un milieu de travail et ne pas confondre évaluation « managériale » et évaluation des acquis d'une formation sont le plus souvent ignorés. Il faudrait adopter le conseil de De Ketele (2010) de ne pas se tromper d'évaluation. Enfin la recherche fréquente d'associer AFEST et certification, donc de raisonner et de construire une AFEST compatible avec une certification existante,

engendre encore d'autres difficultés. Les intervenants chargés d'aider les entreprises ou institutions à mettre en place des AFEST ne sont pas toujours conscients de ces aspects du dispositif et de la place qu'y occupent ou devraient occuper les actions sur différents types d'évaluation.

L'objet de ce texte va donc être de développer rapidement les différents défis évaluatifs en jeu dans la conception et la conduite d'une AFEST, un peu plus largement, d'une formation en milieu de travail, et d'apporter quelques éléments d'outillage conceptuel et méthodologique pour répondre à ces défis. Il vise, peut-être avant tout, à mettre en évidence la nécessité de penser ces formations du point de vue évaluatif et dans une logique d'évaluation formative.

III. Les conditions « institutionnelles » de l'évaluation

L'évaluation en milieu de travail, notamment pour des salariés, est une activité sensible. La mise en question des compétences peut-être une source de conflit, mais d'abord une source de déstabilisation d'un salarié. Si l'employeur ou le responsable hiérarchique sont en position pour traiter des compétences et des moyens de faire évoluer les compétences d'un salarié, en particulier par le biais des entretiens professionnels, leur entrée dans ce qu'on peut appeler l'espace de la formation risque d'entraîner des difficultés. Dans un chantier conduit pour une grande entreprise de distribution d'électricité (Mayen et Gagneur, 2017), la recherche avait ainsi mis en évidence, avec les partenaires, la nécessité d'assurer ce qui avait été appelé « une bulle de confidentialité », concernant notamment l'évaluation, pour distinguer l'espace de la formation de l'espace du travail où le rôle du manager s'exerce pleinement. C'est à lui, en relation avec l'agent bénéficiant de la formation, d'évaluer les compétences de ce dernier et d'en tirer les conséquences en termes de distribution du travail ainsi que de proposer des formations quand cela s'avère nécessaire. En revanche, les « moments » évaluatifs, positionnement avant ou en début de formation, évaluation en cours de formation et évaluation finale, relèvent d'un cadre et d'acteurs distincts des « managers ».

Ce qui est facile à mettre en place dans une grande structure ne l'est pas lorsqu'il s'agit d'une petite. Pourtant, il est possible de dissocier espace et temps de la formation de ceux du travail. Le temps de la formation est d'abord celui de la mise en jeu de ce qu'on sait faire ou non, de la confrontation à des situations et tâches pas ou moins familières et

maîtrisées, le temps du doute, de l'essai, des erreurs, des hésitations, des remises en cause. C'est aussi ce que la recherche déjà citée avait mis en lumière, la nécessité d'un espace et d'un temps sécurisés hors du regard et des jugements de la hiérarchie et des collègues, de droit au doute, à l'hésitation, au temps de comprendre et d'essayer, à l'erreur... La sécurité est une condition d'apprentissage fondamentale. On observe, dans le monde de la santé, par exemple, les difficultés rencontrées par les élèves en stage pour pouvoir apprendre avec ces conditions de doute autorisé, de temps pour apprendre et comprendre, de sécurisation. Ils ne peuvent pas toujours trouver les moyens de mettre en place ce qu'ils ont appris tant ils sont sous le regard urgent de collègues débordés par leur propre travail.

IV. Analyse du travail et analyse de la formation : un processus de référentialisation

La notion de « référentialisation », telle que Gérard Figari (1994) et Figari et Remaud (2014) l'a développée, est très fructueuse pour raisonner non seulement les dimensions évaluatives d'une action de formation, mais aussi pour l'ensemble de celles-ci. L'identification des besoins de formation relève d'emblée du processus de construction des savoirs et savoir-faire (des compétences) de référence, ceux qui vont faire l'objet de la formation, ceux qui seront évalués, ceux qui pourront être reconnus et certifiés. En didactique professionnelle, l'analyse des besoins commence, certes, par la prise en compte de la demande de formation énoncée par une organisation, mais, très vite, opère un retour vers le travail : quelles sont les situations et tâches à maîtriser, quelles en sont les caractéristiques, en termes de complexité, de risques, de variations des conditions, de gestes professionnels, de raisonnements, d'enjeux productifs, d'évolutions actuelles ou à venir d'une part, en termes de difficultés, erreurs, incidents, durée et obstacles d'apprentissage rencontrés par les salariés, autrement dit en termes de compétences d'autre part. Sur ce dernier point, la démarche cherche aussi à identifier, avec des professionnels déjà compétents, les modes de raisonnements et de réalisation de l'action qui répondent aux exigences des situations, afin d'identifier les savoirs et savoir-faire « compétents » qui pourront constituer la référence, et donc les objectifs à atteindre par la formation. Elle s'intéresse aussi au parcours de ceux qui sont devenus « experts ». En effet, leurs compétences se sont développées au long d'un parcours de formation et d'expérience qui peut ou non être « reproduit » pour la formation à mettre en place : le plus fréquemment, on peut constater que les professionnels compétents le sont

devenus avec une certaine durée d'expérience (comment alors la mettre en place ou la compenser par la formation ?) ou bien ont fait certaines expériences qui ont fortement contribué à leur expertise actuelle (qu'il faut là encore pouvoir reproduire ou compenser), ou bien ils ont bénéficié de certaines formations qui entrent aussi dans l'élaboration de leur expertise.

On reconnaît ici les deux versants de la démarche de didactique professionnelle, analyse du travail pour la formation, analyse de la formation (ici du parcours d'expérience des professionnels de référence) pour le travail (Mayen, Pastré et Olry, 2010). On reconnaît aussi des aspects essentiels du processus de référentialisation, principalement la recherche et la combinaison de critères de références issus a) du point de vue de la demande, donc de la hiérarchie qui prescrit et contrôle le travail, b) du travail lui-même, ce que l'ergonomie appelle « la tâche », autrement dit ce qui est à faire, les buts à atteindre dans des conditions déterminées, c) de la manière dont le travail est réalisé par ceux qui le maîtrisent, et par ceux qui ne le maîtrisent pas complètement, d) des aspects « apprentissage » du travail avec ses caractéristiques de complexité, d'habileté, de temps pour l'apprendre, d'obstacles...

La négociation porte aussi sur le niveau de performance attendu et sur le niveau de compétence correspondant. Tout d'abord, il s'agit de partir de l'obligation, pour une AFEST, de ne pas limiter l'apprentissage à la maîtrise d'une situation spécifique, mais d'étendre celle-ci à une famille plus large de situations. Pour cela, on peut recourir à deux notions, celles de diversité et de variabilité. La diversité concerne les caractéristiques des situations d'une même classe. Par exemple, la toilette d'une personne âgée est différente à domicile ou dans une institution, d'un logement à un autre selon l'aménagement plus ou moins favorable du logement, selon l'état de santé et les capacités de la personne. La variabilité concerne les variations, notamment de complexité, au sein de la même situation. Ainsi pour une même personne, l'état de ses capacités physiques, de son moral, de sa santé peut changer d'un jour à l'autre, nécessitant une prise d'information continue, mais aussi une adaptation de l'action d'un ou d'une professionnelle. Il s'agit donc de décider quelle extension de la classe de situations peut maîtriser une personne, à quelles variations et complexités elle peut faire face. Et, par conséquent, quelles connaissances, raisonnements et modes d'action doit-elle construire ou développer pour y parvenir ?

Pour une entreprise ou structure donnée, l'encadrement hiérarchique aura ainsi à décider du degré de variabilité, complexité, diversité, visé, autrement dit, à la fin de la formation la personne pourra maîtriser la

situation spécifique ou une classe plus large, avec tel degré d'autonomie et de maîtrise des variations et donc de la complexité.

Les décisions de cette évaluation de ce qui est visé concernent les moyens à mettre en place, mais posent aussi la question : la ou les situations de travail disponibles ont-elles le potentiel suffisant pour remplir ces objectifs ? A quels aménagements doit-on procéder pour les rendre formatives ? Quelles situations formatives trouver, dans d'autres milieux de travail ou en organisme de formation, pour compléter le parcours ?

Pour terminer, la définition du « référentiel » peut se complexifier encore dans le cas d'une mise en relation avec une certification existante qui peut excéder les objectifs visés par une entreprise ou structure professionnelle. Ce travail de référentialisation est cependant fondamental et il permet de disposer, dès cette phase d'analyse, des critères suffisants pour une évaluation-positionnement préalable, mais aussi pour l'évaluation finale, voire pour la reconnaissance et la certification des acquis.

V. Évaluer le potentiel formatif des situations

La notion de potentiel formatif est issue de l'idée selon laquelle une situation de travail est plus ou moins favorable à l'apprentissage, à la construction et au développement des compétences pour un individu ou un groupe de personnes à un moment donné de leur trajectoire professionnelle. Le potentiel formatif peut ainsi être plus ou moins élevé, voire négatif, lorsqu'il engendre des apprentissages de conduites à risques, de pratiques inefficaces, d'une réduction des capacités, de désapprentissages. L'idée s'oppose à une représentation assez courante selon laquelle toute expérience serait apprenante, et selon laquelle, grosso modo, une situation dénommée à peu près de la même manière d'un lieu de travail à un autre, serait également formative. C'est un des postulats de nombreuses formations par alternance. Or, on constate qu'il n'en est rien et que l'iniquité en découle. Ceci aussi bien pour un salarié qui peut bénéficier de conditions d'apprentissage et de développement favorables que pour un stagiaire ou une personne qui bénéficie d'une AFEST. Le potentiel formatif n'est pas seulement fonction de l'entourage, en particulier d'un tuteur ou d'une équipe de travail qui soutiendrait les activités d'apprentissage d'un nouveau ou moins expérimenté. Il dépend de la nature même du travail : les tâches constitutives d'une situation sont-elles présentes ? La personne a-t-elle la possibilité de s'y confronter ? Les conditions « techniques » sont-elles actuelles ou dépassées ? Certaines

caractéristiques de diversité et variabilités sont-elles présentes ? Ainsi, une expérience d'aide à des personnes âgées en bon état de santé, ne permet-elle pas d'apprendre l'aide à des personnes plus dépendantes. Dans un atelier de réparation automobile, si les tâches complexes sont distribuées aux techniciens les plus compétents, pour des raisons d'efficacité de la répartition du travail et de rentabilité, le « stagiaire » ne peut se confronter à ces situations et les salariés estimés moins compétents n'ont pas l'occasion de réactiver leurs capacités qui se dégradent rapidement.

Dans un article intitulé *s'écarter du travail pour mieux l'apprendre* (Mayen, 2021), nous avions mis en évidence la nécessité d'aménager le travail pour mieux le rendre apprenant, d'une part à partir de ce qui vient d'être développé, d'autre part, pour tenir compte des facteurs d'apprentissage en jeu, par exemple, possibilités d'interrompre le flux du travail pour évaluer l'évolution du processus, prendre conscience, raisonner les liens entre actions et effets de l'action, possibilités d'hésiter, de se tromper, de recommencer, d'essayer, d'accéder à ce qui se passe avant et après son poste de travail, de demander aide et explications, possibilités de faire varier les conditions, etc.

L'activité d'évaluation du potentiel formatif des situations de travail suppose donc une démarche d'analyse du travail et de ses caractéristiques en termes d'apprentissage. Dans la conception d'une AFEST, ce passage obligé passe relativement bien auprès des partenaires, car il est central : que va-t-on effectivement pouvoir faire avec les situations disponibles dans le lieu de travail ? Il semble que cela peut éclairer ceux qui ont à piloter les formations duales ou par alternance par la focalisation qu'il établit sur les situations elles-mêmes et leurs particularités qui sont toujours aussi des particularités formatives.

VI. Evaluer les compétences

Une grande partie de ce qui concerne cette dernière catégorie d'évaluation a déjà été abordée précédemment. Car les différentes étapes des démarches précédentes constituent le système de référence à partir duquel les activités d'évaluation des compétences peuvent être référencées. Ainsi, l'estimation du potentiel d'apprentissage d'une situation est-il en même temps une démarche d'estimation du potentiel d'évaluation des compétences. Les critères que nous avons énoncés pour l'une étant aussi les critères pour l'autre.

Il semble essentiel de penser la mise en place d'un positionnement qui peut se faire de deux manières, par confrontation directe avec l'action en situation ou par analyse anticipatrice de l'action en situation. Le choix de la seconde proposition présente un avantage cognitif et régulateur. Cognitif parce qu'il est demandé à la personne en formation une série d'activités à fort potentiel formatif : remobiliser ses connaissances et expériences sur la situation ; exprimer ce qu'elle sait faire ou pas ; faire un diagnostic de la situation : (Quelle est cette situation ? Quels sont les buts et les résultats attendus ? Quelles sont les conditions ?) ; proposer des décisions d'action et des points d'attention ; exprimer ses doutes, ses hésitations, ses méconnaissances, voire ses craintes. Par la distance qui est prise avec l'action elle-même et par l'obligation d'exprimer pour une autre sa compréhension de la situation et ce qu'il pense devoir faire, par la conscience qu'il doit en prendre, l'apprenant est mis en situation d'effort cognitif exigeant et formateur. Par ailleurs, il s'engage dans un processus d'autoévaluation moteur pour l'engagement dans la formation. De son côté, celui qui occupe la position de formateur peut effectuer une première évaluation de ce que sait et sait faire celui qu'il va accompagner, et ainsi intervenir pour former, aménager la situation, diriger son attention sur certains aspects critiques de l'action de celui qui va apprendre. La première mise en situation va ensuite jouer deux rôles, celui de formation, mais aussi celui de situation de positionnement-évaluation.

Un des outils les plus pertinents qui peut être mobilisé par les formateurs à ce stade est constitué de deux modèles (Mayen et Lainé, 2019) :

Celui, fondamental dans tout apprentissage par expérience : la mise en relation des buts et résultats attendus avec les résultats obtenus ; la mise en relation des résultats obtenus avec le décours des actions réalisées et l'évolution des conditions de l'action.

Celui proposé par Savoyant (2010) qui définit l'action de la manière suivante : toute action est finalisée par des buts et réalisée par une succession d'opérations : opérations d'exécution, autrement dit d'enchaînement d'actions motrices effectrices (ce qu'on fait et ce qu'on dit) et opérations d'orientation, autrement dit, interprétation des buts et résultats attendus, diagnostic des conditions de la situation, élaboration de l'action et des opérations à mettre en œuvre, identification des points de contrôle de l'action, contrôle et régulation de l'action. La réflexivité s'exerce donc conjointement sur les opérations concrètes de réalisation de l'action, mais aussi sur les opérations cognitives de mobilisation de connaissances et de raisonnement. Ce qui correspond aux exigences d'une AFEST. Là encore, ce qu'on nomme réflexivité correspond aussi à une authentique

activité d'autoévaluation pleinement formative, mais à condition d'être ainsi outillée.

VII. Conclusion

Dans le cadre restreint de ce texte, nous avons essayé de mettre en évidence les enjeux évaluatifs des AFEST. Ils sont nombreux et structurels. Cela signifie que ceux qui ont à mettre en place ces actions de formation en situation de travail pourraient utilement les prendre en compte, et adopter aussi un regard évaluatif, pas seulement pour remplir les obligations d'évaluer, mais aussi, plus globalement pour réaliser les opérations d'ingénierie et de conduite des formations. On note que la mobilisation du cadre théorique et méthodologique de la didactique professionnelle, ou plus largement des démarches d'analyse du travail pour la formation et d'analyse de la formation pour le travail, constitue une part essentielle de la démarche. L'analyse des situations et de leurs constituants, notamment en termes d'apprentissage, l'analyse des modes de raisonnement et d'action des professionnels compétents et des moins compétents ainsi que l'analyse du potentiel formatif des situations demandent d'utiliser quelques concepts, méthodes et de porter aussi un certain regard sur ce qu'est le travail, mais aussi ce que sont les conditions d'apprentissage du travail.

L'intérêt et le défi des AFEST amène à repenser les formations duales ou par alternance qu'on « regarde » peut-être trop souvent un peu d'en haut. Une AFEST oblige à concentrer l'attention sur une classe particulière de situations au sein d'un emploi ou métier plus large. On assiste à une mise en place d'AFEST pendant des parcours de formation par alternance. Ce qui pourrait paraître paradoxal. Que ne fait-on le même effort pour toutes les situations de formation par alternance ? Quoi qu'on en pense, cela est sans doute une bonne chose puisque cela peut amener à considérer de plus près le potentiel formatif réel des situations de travail et de l'expérience que des gens peuvent en faire.

Cela ouvre des voies de recherche nouvelles. Les premières, sans doute les plus difficiles à mettre en place, pourraient consister à introduire, dans quelques dispositifs d'AFEST, une série de "moments évaluatifs" et d'examiner leurs effets sur les pratiques des acteurs et sur l'apprentissage des individus. On pourrait faire l'hypothèse que l'introduction d'une problématique générale évaluative, telle que nous en avons suggéré quelques aspects, aiderait les acteurs chargés de l'ingénierie de conception,

et ceux chargés de la conduite de ces formations, à structurer leurs interventions et à réguler leurs pratiques, les aidant ainsi à trouver une ligne directrice qu'ils peinent souvent à découvrir avec les seuls repères du cadre juridique. Un autre type de recherche pourrait examiner davantage le rôle et l'effet de différents types de réflexivité, et notamment de séquences de réflexion avant, pendant et après l'action, et en relation étroite avec le travail. Là encore, deux hypothèses pourraient être examinées concernant un plus grand développement de la part cognitive de l'action, et concernant une plus grande facilité pour les formateurs à amener les apprenants à accéder à une pensée pour l'action.

VIII. Bibliographie

Begon, E. et Duclos L. (dir.). (2021). L'afest, cadrage et débordements [numéro thématique]. *Education Permanente* (227).

De Ketele, J.-M. (2010) Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, *15*(1), 25-37.

Dewey, J. (2012). *Expérience et nature*. Gallimard.

Figari, G. (1994). *Évaluer : quel référentiel ?* De Boeck.

Figari, G. et Remaud, D. (2014). *Méthodologie d'évaluation en éducation et en formation*. De Boeck.

Mayen, P. (2021). S'écarter du travail pour mieux l'apprendre. *Éducation permanente, 1*(226), 51-68.

Mayen, P. et Gagneur, C.-A. (2017). Le potentiel d'apprentissage des situations, une condition pour la conception de formations en situations de travail. *Recherches en Éducation*, 28. https://journals.openedition.org/ree/6050

Mayen, P. et Lainé, A. (2019). *Valoriser le potentiel d'apprentissage des expériences professionnelles.* Educagri Editions. Coll. Praxis.

Mayen, P., Pastré, P. et Olry, P. (2024). Ingénierie didactique professionnelle. Dans P. Carré, P. Caspar, C. Frétigné et O. Las Vergnas (dir.), *Traité des sciences et des techniques de la formation* (p. 564-582). Dunod.

Piaget, J. (1992). *Réussir et comprendre*. Presses Universitaires de France.

Savoyant, A. (2010). Éléments d'un cadre d'analyse de l'activité. *Travail et apprentissages*, (5), 43-54.

Tardif, J., Borges, C. et Malo, A. (dir.) (2012). *Le virage réflexif en éducation*. De Boeck.

Chapitre 17. Quand dire l'évaluation, c'est faire apprendre

La performativité des discours sur l'évaluation

Marc Romainville

I. Introduction

Longtemps, les pratiques d'évaluation des acquis sont restées, dans l'enseignement supérieur, à l'abri de la recherche. Régies par la coutume et l'implicite au temps des héritiers (Bourdieu et Passeron, 1964), ces pratiques ne faisaient l'objet ni d'explicitation systématique ni d'inventaires précis et encore moins d'études scientifiques (Romainville, 2002). Ainsi, la docimologie des années 1980 – science dite « des examens » (De Landsheere, 1992) – s'arrêtait respectueusement aux portes de l'enseignement supérieur.

Une série d'évolutions ont ensuite contribué à progressivement lever le voile sur ces pratiques : épinglons parmi ces évolutions l'arrivée massive de « nouveaux étudiants » – moins enclins à décoder grâce à leur capital culturel des exigences implicites – et la professionnalisation du métier d'enseignant à côté de celui de chercheur notamment. L'organisation – en 2011 et sous l'impulsion de Rémi Goasdoué et Marc Vantourout – du premier colloque de l'ADMEE consacré explicitement à l'enseignement supérieur témoigne de ce changement. L'importante diversité des contributions à ce colloque montre que les pratiques d'évaluation touchent en réalité l'ensemble des dimensions de ce palier du système éducatif et

qu'elles sont loin de se cantonner à l'évaluation des acquis : l'évaluation porte aussi sur la productivité en recherche, sur la qualité des programmes et des établissements et même – *horresco referens* – sur la qualité des enseignements telle qu'elle est perçue par les étudiants (Romainville et al., 2013). La recherche s'empare alors à bras le corps de ces questions.

S'agissant de l'évaluation des acquis, les études se sont développées dans plusieurs directions. Sur un plan descriptif – qui pouvait du reste virer rapidement au normatif –, elles ont d'abord tenté de cartographier les pratiques et d'étudier la manière dont ces dernières rencontrent, ou pas, les qualités traditionnellement attendues de pareilles évaluations (fidélité, validité, transparence et gestion des enjeux sociaux sous-jacents essentiellement). D'un point de vue plus compréhensif ensuite, la recherche a visé à mieux comprendre ce qui incite les enseignants à se tourner vers tel ou tel type de pratique, que leurs motifs soient d'ordre circonstanciel (le niveau d'enseignement, le nombre d'étudiants, etc.) ou représentationnel (leur conception de l'enseignement notamment).

Une perspective de recherche particulièrement novatrice – et qui fera l'objet du présent chapitre – a trait à ce que l'on pourrait appeler la performativité des discours des enseignants sur l'évaluation. On s'aperçoit en effet que l'évaluation ne peut être réduite à un acte pédagogique technique qui consisterait à enregistrer des acquis, mais que les discours des enseignants sur leurs pratiques évaluatives ont des effets, par eux-mêmes, sur les apprentissages des étudiants. Bien qu'il soit l'objet de critiques internes à la linguistique, le concept de performativité présente l'intérêt de souligner que les actes de langage, au-delà de leur fonction de description du réel, produisent des effets par leur seule énonciation (Austin, 1970 ; Marignier, 2021). Appliqué à l'évaluation des acquis, le concept ouvre un champ de recherche particulièrement fécond. Il permet d'envisager l'évaluation, non plus seulement comme une action pédagogique d'appréciation de l'atteinte d'objectifs d'enseignement, mais comme un acte perlocutoire qui crée des changements et directement du « faire » auprès des étudiants. Dans cette perspective, les dires des enseignants sur l'évaluation produisent, par eux-mêmes, des effets sur les pratiques étudiantes. A notamment été exploré l'impact des modalités d'évaluation annoncées sur l'approche des étudiants (en profondeur, en surface ou stratégique), sur leurs croyances épistémiques et sur leur sentiment de « pouvoir agir ». Des perspectives de recherches prometteuses se dessinent à partir des premières études menées sur cette thématique ; ces perspectives seront évoquées en conclusion.

II. Missions, fonctions et effets de l'évaluation

En matière d'évaluation des acquis, il est fécond de distinguer les missions, les fonctions et les effets. Les *missions* renvoient à ce que le système éducatif assigne explicitement comme objectifs à cette évaluation via les textes réglementaires qui la régissent (loi, décret, circulaire, etc.). En ce qui concerne l'évaluation des acquis des étudiants de l'enseignement supérieur, on peut aisément identifier trois missions essentielles qui lui sont traditionnellement confiées : la première – sans doute la principale d'un point de vue institutionnel – vise à certifier l'acquisition par les étudiants des savoirs, savoir-faire et compétences constitutifs des objectifs du programme de formation dans lequel ils sont inscrits et, sur cette base, de délivrer les certifications et les diplômes *ad hoc*. La deuxième mission, assignée plus récemment, est de favoriser les apprentissages des étudiants. C'est la fonction formative de l'évaluation des acquis qui a fait sa joyeuse entrée dans le monde du supérieur, certes timidement et en y rencontrant pas mal d'écueils (Romainville, 2021), mais de manière quand même résolue. Parfois, une troisième mission est également explicite, celle de sélectionner les étudiants (par exemple, à l'occasion des concours d'entrée en spécialités de médecine) et/ou de les classer, y compris en fin de formation, comme jusqu'à très récemment, les classements de sortie en vigueur dans certaines Grandes écoles françaises.

Par analogie à la fonction de reproduction des inégalités sociales dévoilée par Bourdieu et Passeron dans leur ouvrage de 1970, on parlera de *fonction*s pour désigner des résultats décelés à partir d'une analyse du fonctionnement des pratiques d'évaluation des acquis, qui répondent à un besoin ou à une fin, mais sans que ce besoin ou cette fin ne soit ni revendiqué ni assumé comme tel. Ainsi, même lorsqu'elle n'apparaît pas dans les textes réglementaires, la fonction de sélection des étudiants peut être considérée comme étant à la source de certaines pratiques d'évaluation telles que la normalisation des notes permettant d'obtenir un taux de passage du premier au deuxième cycle universitaire jugé tacitement acceptable. Pour en revenir à Bourdieu et Passeron, ils avaient d'ailleurs montré, précédemment à leur ouvrage de 1970, que les exigences « diffuses et implicites » de l'évaluation avaient pour fonction, dès lors qu'elles étaient plus facilement décodables par les héritiers qui bénéficiaient d'un capital culturel important pour ce faire, de contribuer à la reproduction des inégalités sociales (Bourdieu et Passeron, 1964). Et la persistance de cette fonction continue d'être observée des décennies après

l'ouvrage de Bourdieu et Passeron comme le montre, par exemple, l'étude de Williams (2005).

Enfin, les pratiques d'évaluation des acquis produisent aussi des *effets*, « collatéraux » pourrait-on dire, qui ne se réduisent ni aux buts assignés via les missions, ni aux fonctions tacites que cette évaluation exerce. Par le fait de se présenter de telle ou telle manière ou de se pratiquer selon telle ou telle modalité, l'évaluation des acquis entraîne des conséquences qui ne sont *a priori* ni recherchées par quiconque (que ce soit explicitement en tant que mission ou implicitement en tant que fonction), ni dès lors anticipées. C'est l'un de ces effets sur l'apprentissage des étudiants que le présent chapitre entend explorer.

III. L'évaluation comme acte perlocutoire

Dans le registre des effets « collatéraux » des discours sur l'évaluation, il en est un qui est assez remarquable du fait qu'il constitue une répercussion en amont puisqu'il engendre des conséquences avant que l'objet du discours – l'acte évaluatif – n'ait eu lieu (Entwistle et Entwistle, 1991). De nombreuses recherches ont en effet montré que les étudiants, durant leurs apprentissages et donc avant que l'évaluation de ces derniers ne se déroule, sont véritablement pilotés par les modalités d'évaluation qu'ils anticipent et à propos desquelles ils recherchent frénétiquement des informations auprès des pairs des années ultérieures notamment. Autrement dit, ils adaptent, plus ou moins consciemment, leur approche et leurs stratégies à la manière dont ils perçoivent que l'évaluation de la maîtrise des contenus de cours sera organisée. Comme nous le verrons plus loin, cet effet retour s'exerce aussi sur d'autres dimensions de l'apprentissage telles que les croyances épistémiques, l'ajustement académique ou le sentiment de « pouvoir agir ».

C'est dans ce sens que l'on pourrait parler, par analogie à un cadre théorique de la linguistique évoqué plus haut, de performativité des discours tenus, explicitement ou implicitement, par les enseignants à propos des modalités d'évaluation des acquis. Ces discours, par le simple fait qu'ils sont tenus par les enseignants ou par d'autres acteurs – notamment les anciens étudiants qui sont déjà passés par là et qui savent « à quelle sauce ils ont été mangés » – et qu'ils sont interprétés par les étudiants, produisent des changements dans leurs pratiques d'apprentissage et, de manière plus générale, dans leur rapport au savoir.

Toujours en s'inspirant de concepts de la linguistique, on pourrait donc considérer que les discours tenus par les enseignants sur les modalités d'évaluation constituent des actes perlocutoires à rebours (Oswald, 2020). En effet, au-delà de leur intention première de communiquer – à la suite d'une exigence nouvelle de transparence – de manière descriptive la manière dont l'évaluation se déroulera (fonction locutoire), ces discours auront aussi une portée sociale via les effets (perlocutoires) qu'ils produiront sur les actes en amont de leurs destinataires. Ce qui est intéressant dans la distinction locutoire/perlocutoire, c'est le fait que si les locuteurs contrôlent assez bien la fonction locutoire et descriptive de leur discours, les effets perlocutoires échappent grandement à leur contrôle (Oswald, 2020). Ainsi, un discours peut entraîner accidentellement des conséquences que le locuteur n'avait pas recherchées, tout comme il se peut aussi qu'un locuteur ne parvienne pas à ce que les étudiants s'engagent dans des comportements conformes à l'intention de son discours. Par exemple, c'est le cas d'enseignants qui annoncent et mettent en place une évaluation des acquis réalisée via un portfolio dans le but de favoriser une approche en profondeur chez les étudiants, mais qui se disent dépités de constater que ces derniers ont plutôt opté pour une approche stratégique, leur réception du discours enseignant ayant été différente de ce qui avait été anticipé (Romainville et Fischer, 2020). Nous reviendrons sur cet aspect, mais soulignons dès à présent que ce qui produit des effets perlocutoires ne réside donc pas tant dans le discours des enseignants que dans la compréhension que les étudiants s'en construisent.

IV. Des effets perlocutoires variés et bien établis

Des effets perlocutoires ont été mis au jour à propos de plusieurs dimensions de l'apprentissage des étudiants. À tout seigneur tout honneur, il a été montré que les discours tenus sur l'évaluation des acquis influencent l'approche d'apprentissage développée par les étudiants. Conceptualisée dès les années 1970 et confirmée ensuite par de nombreuses recherches, une typologie parmi les plus célèbres en matière d'apprentissage étudiant distingue trois types de rapport aux études supérieures : l'approche en surface – l'étudiant cherche avant tout à reproduire fidèlement un discours en le mémorisant tel quel et en le renvoyant à son producteur –, l'approche en profondeur – mû par des enjeux personnels de recherche de sens, l'étudiant s'efforce de comprendre ce qu'il y a derrière les mots et les textes – et l'approche

stratégique – l'étudiant cherche avant tout à réussir et consacre son énergie à gérer de manière optimale les ressources à sa disposition (Cuthbert, 2005 ; Marton et Säljö, 1984).

Cette typologie ayant connu une large diffusion internationale, de très nombreuses études l'ont incluse dans leur méthodologie et l'on dispose dès lors d'une vue d'ensemble assez solide des effets perlocutoires de l'anticipation des conditions d'évaluation sur l'approche privilégiée par les étudiants (Cope, et Staehr, 2005 ; Dahlgren et al., 2009 ; Entwistle et Entwistle, 1991 ; Entwistle et Karagionnopoulou, 2014 ; Hailikari et al., 2022 ; Lizzio et al., 2002 ; Scouller, 1998 ; Struyven et al., 2005).

Schématiquement, les étudiants auront tendance à mettre en place une approche en surface s'ils perçoivent – le fait que ce soit à tort ou à raison importe peu puisque ce sont les interprétations des étudiants qui comptent et non le contenu objectif des discours – a) que l'évaluation de leurs acquis sera centrée sur la restitution de faits ponctuels et isolés – le Questionnaire à Choix Multiple (QCM) et pire encore le « texte à trous », archétypes emblématiques d'une telle évaluation, enclenchent des effets perlocutoires particulièrement malheureux – ; b) qu'ils seront confrontés à des exigences « diffuses et implicites », arbitrairement fixées par l'enseignant et sur lesquelles ils n'auront aucune prise ; c) que l'aspect quantitatif de la performance attendue (nombre d'éléments corrects, nombre d'exercices ou de parties d'exercice résolus, etc…) et de son appréciation – sous la forme de notes chiffrées, décomposées en sous-notes justificatives – prendra le pas sur un aspect plus qualitatif ; d) que la charge de travail requise pour rencontrer les exigences de l'évaluation sera excessive et démesurée.

A contrario, il y a davantage de chances que les étudiants privilégient une approche en profondeur s'ils estiment a) que l'évaluation portera sur des niveaux taxonomiques autres que la restitution (compréhension, application, analyse ou jugement) présentant davantage de « challenge » intellectuel et qu'elle sera réalisée via des modalités flexibles et des consignes et questions ouvertes nécessitant des textes longs, voire des essais (les outils tels que l'examen à « livre ouvert », le journal d'apprentissage ou le portfolio étant des supports emblématiques de pareilles évaluations, tout comme bon nombre de formes d'évaluation continue) ; b) que le contrat didactique est clair et transparent et qu'aucune « surprise » ne risque d'advenir au moment de l'évaluation par rapport à ce qui a été travaillé durant la formation ; c) que l'aspect qualitatif de l'évaluation primera en matière d'exigences et de retours sur leur performance (par exemple, via un feedback centré sur des commentaires qualitatifs et un jugement global exprimé sobrement en termes de maîtrise,

ou non, des objectifs d'enseignement) ; d) que la charge de travail requise sera raisonnable, ce qui les autorisera à développer une réflexion personnelle, voire critique sur les contenus à apprendre.

De manière à permettre au lecteur de prendre conscience de la force du phénomène perlocutoire, examinons plus en détail l'une des études qui ont permis de dresser ce tableau général, même s'il est un tantinet idyllique dès lors que les effets perlocutoires mis au jour n'apparaissent que sous certaines conditions, comme indiqué plus loin.

Dans un article de 2009, Dahlgren et ses collaborateurs rapportent un effet étonnant, étant donné qu'il prend sa source dans une différence apparemment anodine entre deux systèmes d'évaluation. Dans le premier – très classique en Suède avant l'harmonisation impulsée par le processus de Bologne –, les étudiants étaient évalués selon un jugement binaire « réussite/échec » : les objectifs de l'enseignement étaient atteints ou pas, point à la ligne. Pour des raisons d'harmonisation européenne, certains programmes suédois sont passés à un système de notation en 7 degrés (2 pour l'échec et 5 pour la réussite). 402 étudiants (125 enrôlés dans des programmes « réussite/échec » et 277 dans des programmes convertis aux 7 échelons) sont interrogés sur leur perception des exigences des évaluations en termes, soit de reproduction (assurée par une approche en surface), soit de production (nécessitant une approche en profondeur). Les effets de la perception de cette subtile différence entre les deux échelles étonnent par leur ampleur : la proportion d'étudiants qui déclarent une prédilection pour la reproduction en surface est deux fois plus élevée chez ceux qui seront jugés selon une échelle à 7 niveaux par rapport à ceux des programmes « réussite/échec ». Autrement dit, une modalité du système de notation, à elle seule, a incité des étudiants à développer une approche d'apprentissage particulière : soumis à une échelle complexe, ils se sont davantage concentrés sur la tâche d'évaluation en elle-même lors de leur étude que sur l'apprentissage. Toute se passe comme si l'annonce d'une évaluation plus finement graduée avait incité les étudiants à mettre l'accent sur des objectifs reproductifs, en se concentrant non seulement sur les parties essentielles des cours, mais aussi sur des parties moins importantes et en se renseignant sur les évaluations antérieures de manière à tenter de grappiller des points à gauche et à droite, pour « monter » dans l'échelle. Et l'on comprend en effet que l'accès à des échelons plus élevés de la réussite se jouera sur des détails, et notamment sur une accumulation, d'apparence académique, de faits, de dates, de données, fussent-elles assez secondaires par rapport à ce que l'on peut en tirer comme leçons essentielles.

Contextuelle et ne constituant donc pas un trait stable de l'étudiant, l'approche d'apprentissage doit être incarnée, dans une situation donnée, par des stratégies concrètes. Des études, parfois les mêmes que celles évoquées précédemment, ont investigué le lien entre ces stratégies d'apprentissage et les modalités d'évaluation. Elles confirment, de manière générale mais en détaillant plus finement les pratiques d'étude des étudiants, le panorama global auquel les recherches portant sur les approches ont conduit. Ainsi, Barbeau s'est intéressée aux modalités d'évaluation et à leurs effets en regard des préférences des étudiants en la matière (Barbeau, 2025 ; Barbeau et al., 2025). Elle montre notamment que les étudiants qui apprécient les QCM annoncés sont moins enclins à mettre en place des stratégies métacognitives, sans doute parce qu'ils estiment que l'auto-évaluation de leur état de connaissance importera peu par rapport aux exigences de mémorisation de ce type d'évaluation. Au contraire, ils investissent fortement les stratégies de gestion des ressources telles que l'organisation de leurs notes, le classement de leurs documents et l'utilisation d'onglets autocollants.

Dans le même sens, Stanger-Hall (2012) rapporte que les stratégies d'étude actives et critiques (élaborer des synthèses personnelles, réorganiser les informations transmises, rédiger des fiches de comparaison, analyser la manière dont les différentes informations sont reliées entre elles et leur impact sur la vie et la société) sont nettement moins rapportées par les étudiants auxquels on a annoncé une évaluation uniquement réalisée par QCM, malgré les encouragements de leurs enseignants à mettre en œuvre de telles stratégies. Notons toutefois que cet effet disparaît lorsque l'évaluation combine QCM et questions ouvertes. C'est donc un mode fermé de questionnement utilisé seul qui produit des effets négatifs ; nous y reviendrons.

Les effets perlocutoires des discours sur l'évaluation ne concernent pas que le rapport à l'apprentissage des étudiants ; ils s'étendent à leur rapport au savoir lui-même. Ainsi, selon de nombreuses études, les croyances épistémiques des étudiants sont aussi impactées par ce qu'ils pensent être les modalités d'évaluation de leurs acquis. On désigne par « croyances épistémiques » les représentations naïves que les étudiants se construisent à propos de la nature du savoir scientifique et des processus de son élaboration (pour une synthèse en français, voir Crahay et Fagnant, 2007). En particulier, un des facteurs d'intégration académique des étudiants au monde de l'enseignement supérieur réside dans l'évolution progressive de leurs croyances épistémiques. D'une position de départ absolutiste (une connaissance scientifique est définitivement vraie ou fausse et il est possible d'arriver à une telle connaissance pour toute question que l'on

peut se poser sur le monde), ils doivent progressivement endosser des croyances davantage relativistes telles que celles qui président au fonctionnement des sciences modernes : toute connaissance est relative au modèle dans le cadre duquel elle a été élaborée et elle est donc appelée à évoluer (Hofer et Pintrich, 1997).

Les effets perlocutoires des discours relatifs à l'évaluation sont, ici aussi, assez impressionnants (Entwistle et Entwistle, 1991 ; Nieminen et Lahdenperä, 2021 ; Sheppard et Gilbert, 1991 ; Tolhurst, 2007). Sans crier au loup en faisant du QCM un bouc émissaire qui ne le mérite pas entièrement, on peut toutefois comprendre comment des croyances absolutistes peuvent être consolidées par cette pratique. En effet, un concepteur de QCM – qui s'est souvent rabattu « à l'insu de son plein gré » sur cette modalité économe, accablé par un très grand nombre d'étudiants à évaluer – prendra un soin particulier à rédiger a) une amorce de question claire, précise et dénuée de toute ambiguïté ; b) une « bonne » réponse entièrement « juste » et non discutable ; c) des distracteurs certes crédibles mais entièrement faux et qui ne contiennent donc aucune once de vérité. De plus, sa quête de décidabilité incitera le concepteur à se tourner vers des sujets ponctuels, circonscrits, isolés, fuyant comme la peste les situations complexes qui ne pourraient pas être tranchées en noir ou blanc. Finalement, les étudiants qui seraient confrontés régulièrement à de pareilles questions d'évaluation finiraient par se convaincre que toute valeur de vérité sur le monde peut en effet être résolue de manière dichotomique et définitive : elle est soit vraie, soit fausse, sans faux-fuyant possible.

L'évaluation influence aussi la manière dont les étudiants se représentent les constituants épistémologiques de disciplines particulières (Nieminen et Lahdenperä, 2024). Ainsi, de nombreux étudiants en mathématiques estiment que l'examen traditionnel est un outil correct et acceptable de mesure de leurs acquis, dès lors que cette discipline est « objective » et que l'existence d'une pluralité de démarches de résolution de problème n'empêche pas qu'une seule d'entre elles aboutisse à la « bonne réponse ». Ils renforcent alors leurs croyances épistémiques absolutistes et leur vision positiviste et « descendante » du savoir dont la maîtrise est uniquement évaluable, à leurs yeux, par un enseignant expert. En corollaire, ils rechignent à accorder de la valeur à des formes d'évaluation qui feraient appel à leur propre appréciation (auto-évaluation) ou à celle d'autres apprenants (évaluation par les pairs).

Signalons enfin que des effets perlocutoires ont été plus récemment investigués à propos d'autres dimensions de l'apprentissage. Par exemple,

Wanner et ses collaborateurs (2024) ont montré que l'annonce d'une plus grande flexibilité dans l'évaluation (choix des modalités, de leurs poids relatifs dans la note finale, des dates de soumission et du type de feedbacks) contribue au développement d'un sentiment de « pouvoir agir » du moins sur les processus d'apprentissage, ce qui n'implique pas, aux yeux des étudiants, une meilleure mainmise sur les résultats (ce que l'étude confirme d'ailleurs).

V. Des réalités plus complexes que les modèles

Si l'existence d'effets perlocutoires des discours sur l'évaluation est solidement établie et que des tendances nettes se dessinent – en particulier, un lien avéré entre, d'une part, l'apprentissage en profondeur, personnel et critique et, d'autre part, des modalités ouvertes d'évaluation –, les études montrent aussi que les situations particulières sont souvent bien plus complexes que les modèles généraux.

Rappelons d'abord que ce ne sont pas les discours sur l'évaluation, par eux-mêmes, qui ont des effets perlocutoires, mais bien les perceptions que les étudiants en ont, et donc la compréhension qu'ils s'en construisent. Entre, d'une part, les intentions des enseignants lorsqu'ils mettent en place et présentent tel ou tel dispositif d'évaluation, et, d'autre part, ce que les étudiants en comprennent ou veulent en comprendre, il y a souvent loin de la coupe aux lèvres. Ainsi, un portfolio présenté comme outil de développement personnel, via l'écriture réflexive et libre, peut être appréhendé, par certains étudiants, comme un instrument de reddition de comptes qui dissimule ses intentions réelles. Ces mêmes étudiants en veulent pour preuve l'ambiguïté du rôle du portfolio dans la certification (Romainville et Fischer, 2020).

De plus, un même discours sur un dispositif sera interprété très différemment par les étudiants en fonction de leurs préférences personnelles. Ainsi, confrontés à un QCM identique, ce sont les étudiants qui ont anticipé des exigences relevant de niveaux taxonomiques élevés qui déclarent davantage privilégier une approche en profondeur (Scouller et Prosser (1994), cité par Scouller, 1998). On a aussi observé que certains étudiants persistent à conserver une approche en profondeur même s'ils savent qu'ils seront évalués par QCM, parce que renoncer à cette tendance personnelle est au-dessus de leur force. Ils payeront d'ailleurs le prix de cette résistance obstinée en obtenant de moins bons résultats (Scouller, 1998).

Il s'agit donc de phénomènes systémiques, au sein desquels il est délicat, voire impossible – et d'ailleurs serait-ce utile ? – de résoudre le paradoxe de la poule et de l'œuf. Ainsi, est-ce parce qu'ils sont confrontés à des QCM que des étudiants développent une approche en surface ou est-ce que leur prédilection pour cette approche les a incités à penser que tout QCM ne peut qu'appeler ce type d'approche ? En réalité, il s'agit sans doute d'une causalité circulaire dans laquelle chaque effet affecte sa propre cause.

Par ailleurs, la « lecture » par les étudiants des discours sur l'évaluation est biaisée par d'autres caractéristiques, ce qui rend particulièrement incertaine la prédiction de leurs effets perlocutoires réels. Pour en revenir à l'exemple du portfolio, des étudiants peuvent ne pas l'interpréter comme une incitation à l'écriture réflexive, libre et personnelle, parce que, dans leur conception de l'apprentissage, l'écriture sur soi n'est pas un moyen efficace pour apprendre ou parce que leur sentiment de compétence en matière d'écriture est faible (Romainville et Fischer, 2020).

Tout ceci explique que de nombreuses données issues d'études empiriques nuancent considérablement les tendances générales modélisées et contribuent ainsi à les affiner.

VI. Perspectives de recherche

À l'avenir, les recherches devraient donc se donner comme objectif de mieux comprendre ces exceptions et leurs logiques, de manière à nuancer et enrichir les modèles généraux. Il s'agirait d'abord de décrire et cartographier plus finement les discours ainsi que les pratiques. Par exemple, les enseignants prennent-ils la peine, dans leurs discours, d'expliciter, au-delà du dispositif, sa logique et son bien-fondé, en regard des objectifs pédagogiques ? Ainsi, on sait que les effets délétères des QCM sur les approches et stratégies d'apprentissage s'atténuent considérablement lorsqu'ils sont utilisés au sein d'un panel d'outils d'évaluation et lorsque l'on a justifié leur présence en regard d'objectifs de mesure précis (Stanger-Hall, 2012). On a aussi observé qu'un QCM plus sophistiqué et visant à mesurer des niveaux taxonomiques dépassant la restitution était de nature à convaincre les étudiants que la compréhension est aussi requise, en complément à la mémorisation, pour la réussite d'une pareille épreuve (Reinmann et Xu, 2005).

Il s'agirait ensuite de mieux comprendre ce qui oriente la « lecture », par les étudiants, des discours sur l'évaluation en investiguant comment leurs préférences et leurs caractéristiques personnelles les incitent à s'en

construire une compréhension particulière, parfois fort éloignée du sens que les enseignants y avaient placé. On retrouve ici une orientation générale que les études sur l'évaluation devraient épouser dans les prochaines décennies, à savoir minorer leur penchant prescriptif et normatif et s'attacher au contraire à étudier les pratiques ordinaires et, pour en comprendre la logique, recueillir de l'information auprès de l'ensemble des protagonistes impliqués dans ces pratiques, y compris leurs destinataires, élèves et étudiants (Bart, 2023).

VII. Bibliographie

Austin, J. L. (1970). *Quand dire, c'est faire.* Éditions du Seuil.

Barbeau, N. (2025). *Rôle de l'évaluation des apprentissages sur la relation entre les stratégies d'apprentissage et l'ajustement académique en contexte de préparation d'examen à l'université.* [Thèse de doctorat, Université Laval].

Barbeau, N., Frenette, E. et Hébert, M.-H. (2025). Effects of perceptions of assessment on the relationship between learning strategies and academic adjustment in a summative exam preparation context. *Social Sciences & Humanities Open, 11*. https://doi.org/10.1016/j.ssaho.2025.101448

Bart, D. (2023). *Évaluation et didactique : un dialogue critique.* Peter Lang.

Bourdieu, P. et Passeron, J.C. (1964). *Les héritiers. Les étudiants et la culture.* Les Éditions de Minuit.

Bourdieu, P. et Passeron, J.C. (1970). *La Reproduction. Éléments d'une théorie du système d'enseignement.* Les Éditions de Minuit.

Cope, C. et Staehr, L. (2005). Improving students' learning approaches through intervention in an information systems learning environment. *Studies in Higher Education, 30*(2), 181–197. https://doi.org/10.1080/03075070500043275

Crahay, M. et Fagnant, A. (2007), À propos de l'épistémologie personnelle : un état des recherches anglosaxonnes. *Revue française de pédagogie, 161*, 79-117. https://doi.org/10.4000/rfp.830

Cuthbert, P. (2005). The student learning process. *Teaching in higher education, 10* (2), 235-249.

Dahlgren, L. O., Fejes, A., Abrandt-Dahlgren, M. et Trowald, N. (2009). Grading systems, features of assessment and students' approaches to learning. *Teaching in Higher Education, 14*(2), 185–194. https://doi.org/10.1080/13562510902757260

De Landsheere, G. (1992). *Évaluation continue et examens. Précis de docimologie*. Labor.

Entwistle, N. et Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: the student experience and its implications. *Higher Education, 22,* 205–227. https://doi.org/10.1007/BF00132288

Entwistle N. et Karagionnopoulou E. (2014). Perceptions of assessment and their influences on learning. Dans C. Kreber, C. Anderson, J. MacArthur et N. Entwistle (dir.), *Advances and innovations in university assessment and feedback* (p. 75-98). Edinburgh University Press.

Hailikari, T., Virtanen, V., Vesalainen, M. et Postareff, L. (2022). Student perspectives on how different elements of constructive alignment support active learning. *Active Learning in Higher Education*, *23*(3), 217-231. https://doi.org/10.1177/1469787421989160

Hofer, B. et Pintrich, P. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88–140. doi: 10.3102/00346543067001088

Lizzio A., Wilson K. et Simons R. (2002). University students' perceptions of the learning environment and academic outcomes. *Studies in Higher Education, 27*(1), 27-52.

Marignier, N. (2021). Performativité. *Langage et société*, hors-série (HS1), 263-266. https://doi.org/10.3917/ls.hs01.0264

Marton, F. et Säljö, R. (1984). Approaches to learning. Dans F. Marton, D. Hounsell et N. Entwistle (dir.), *The experience of learning* (p. 36-55). Scottish Academic Press.

Nieminen, J.-H. et Lahdenperä, J. (2021). Assessment and epistemic (in)justice: how assessment produces knowledge and knowers. *Teaching in Higher Education, 29*(1), 300–317.

Oswald, S. (2020). Pragmatique cognitive, argumentation et perlocution, *Argumentation et analyse du discours* [En ligne], 25. https://doi.org/10.4000/aad.4793

Romainville, M. (2002). *L'évaluation des acquis des étudiants dans l'enseignement universitaire.* Haut Conseil de l'évaluation de l'école.

Romainville, M. (2021). Ce n'est pas en pesant le cochon qu'on le fait grossir ou pourquoi l'évaluation formative peine à se faire une place au soleil. Dans C. Barroso da Costa, D. Leduc et I. Nizet (dir.), *40 ans de mesure et d'évaluation* (p. 81-94). Presses de l'université du Québec.

Romainville, M., Goasdoué, R. et Vantourout, M. (2013). *Évaluation et enseignement supérieur.* De Boeck.

Romainville, M. et Fischer, L. (2020). Dilemmes relatifs à l'usage du portfolio en enseignement supérieur. *Évaluer. Journal international de recherche en éducation et formation, 6*(1), 79-90. https://doi.org/10.48782/x788mg30

Reimann, N. et Xu R. (2005). *Introducing multiple choice alongside short answer questions into the end-of-year examination: the impact on student learning in first year economics* [Conférence]. European Association for Research on Learning and Instruction (EARLI) 11th Biennial Conference, Nicosia, Cyprus, 23 - 27 August.

Sheppard, C. et Gilbert, J. (1991). Course Design, Teaching Method and Student Epistemology. *Higher Education, 22*(3), 229–249. http://www.jstor.org/stable/3447174

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education 35*, 453–472. https://doi.org/10.1023/A:1003196224280

Stanger-Hall, K. F. (2012). Multiple-Choice Exams: An Obstacle for Higher-Level Thinking in Introductory Science Classes. *CBE—Life Sciences Education, 11*(3), 294-306. https://doi.org/10.1187/cbe.11-11-0100

Struyven, K., Dochy, F. et Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment and Evaluation in Higher Education, 30*(4), 325-341. http://www.tandf.co.uk/journals/titles/02602938.asp

Tolhurst, D. (2007). The influence of learning environments on students' epistemological beliefs and learning outcomes. *Teaching in Higher education*, *12*(2), 218-233.

Wanner, T., Palmer, E. et Palmer, D. (2024). Flexible assessment and student empowerment: advantages and disadvantages – research from an Australian university. *Teaching in Higher Education*, *29*(2), 349–365.

Williams K. (2005). Lecturer and first year student (mis)understanding of assessment task verbs: « Mind the gap ». *Teaching in higher education, 10* (2), 157-173.

Chapitre 18. L'évolution du concept de feedback, au carrefour des dimensions cognitive, émotionnelle, comportementale et sociale de l'apprentissage

Des dimensions inhérentes à l'engagement des apprenants dans le traitement du feedback

Matthieu Hausman, Pascal Detroz et Laurent Leduc

I. Introduction

Dans le champ de l'éducation, les feedbacks semblent depuis longtemps pris dans une tension entre leur rôle largement reconnu de soutien aux apprentissages (Hattie et Timperley, 2007) et les limitations qui lui sont imputées par ailleurs (Boud et Molloy, 2013). Cette tension est révélatrice du fait que les feedbacks sont bien en mesure de produire des effets hautement bénéfiques, si les conditions de leur exploitation effective sont réunies, ce qui, en réalité, n'est pas toujours le cas (Sadler, 2010).

Dès lors se posent des questions comme celle de savoir quelles pratiques soutiennent l'efficacité des feedbacks. Quels aspects de la performance ou de la compréhension des apprenants doivent-ils ou ne doivent-ils pas aborder ? À quel point doivent-ils être précis, personnalisés, fréquents ?

Selon quelles modalités les délivrer à leurs destinataires ? Toutes ces questions ont fait l'objet de plusieurs décennies de travaux scientifiques, aboutissant notamment à des références incontournables dans la littérature telles que les méta-analyses de Kluger et DeNisi (1996), Hattie et Timperley (2007) ou encore les revues systématiques de Shute (2008) ou de Brookhart (2008).

Cela étant, la question de l'efficacité des feedbacks demeure complexe, car ils font encore et toujours l'objet d'évaluations peu satisfaisantes dans les enquêtes menées auprès des étudiants, en particulier dans l'enseignement supérieur (Beaumont et al., 2011 ; Pitt et Norton, 2017). De manière concomitante, ils souffrent également d'un manque d'engagement de la part des étudiants dans la recherche, la réception et l'utilisation d'informations utiles pour qu'ils puissent s'autoréguler (Price et al., 2011). En conséquence, nous devons constater que la littérature scientifique recèle encore de nombreuses questions quant à la manière dont les enseignants sont amenés à proposer un retour d'informations à leurs élèves ou à leurs étudiants, afin d'en optimiser les effets sur l'apprentissage.

Depuis quelques années, un courant de recherche s'est penché sur la possibilité d'une autre compréhension du concept de feedback, centrée davantage sur l'apprenant que sur l'enseignant et l'information qu'il est susceptible de fournir au premier. L'intérêt de cette perspective est de favoriser la compréhension des conditions qui soutiennent l'engagement d'actions de la part de l'apprenant, à partir d'un ou plusieurs feedbacks. Ce courant s'est incarné dans une modélisation, de prime abord relativement simple, proposée par Carless et Boud en 2018 sous l'appellation « Student Feedback Literacy ». Ce modèle conceptuel a suscité l'intérêt de la communauté scientifique – non sans alimenter de vifs débats – car il met en évidence les connaissances et aptitudes jugées essentielles à l'engagement des apprenants vis-à-vis des feedbacks. En d'autres termes, il porte sur les principales capacités à développer chez les apprenants pour que ceux-ci tirent le meilleur parti des feedbacks qu'ils reçoivent, mais également pour qu'ils puissent déployer des stratégies proactives par rapport à la réception effective des feedbacks externes. Cette proactivité peut consister en la sollicitation de feedbacks externes auprès de sources pertinentes, ou encore en l'élaboration de feedbacks internes, suffisamment fiables que pour s'appuyer dessus afin de s'ajuster aux besoins identifiés.

Dans le cadre de cette contribution, notre ambition est de revenir sur les conditions d'émergence de cette approche récente des feedbacks dans le

domaine de l'éducation, d'en spécifier quelque peu les composantes avant de nuancer l'importance des aptitudes qu'elles mobilisent, par rapport à la valeur de l'information qui constitue encore et toujours le point de départ des feedbacks.

II. Pourquoi fournir des feedbacks aux apprenants ?

D'après Detroz et ses collègues (2020), les feedbacks destinés aux apprenants s'inscrivent dans une chaîne d'actions propres à soutenir la qualité du processus évaluatif, au même titre que l'analyse du référentiel d'évaluation ou que l'élaboration de l'artefact soutenant la prise de données chez les évalués. En fonction des objectifs poursuivis par la démarche d'évaluation, le feedback sera plus ou moins orienté vers la régulation de l'apprentissage de l'apprenant. Il peut ainsi attester d'un niveau de maîtrise ou de compréhension de contenus donnés par l'apprenant, en tant que feedback de vérification, tout comme il peut susciter une réflexion approfondie chez ce même apprenant, en mettant en évidence des aspects problématiques ou insuffisants vis-à-vis de l'atteinte des objectifs d'apprentissage définis, en qualité de feedback d'élaboration (Kulhavy et Stock, 1989). On comprend donc que la propension des feedbacks à soutenir la régulation de l'apprentissage des apprenants est liée à l'intention pédagogique incarnée par le type d'évaluation auquel sont soumis ces derniers.

Dans le cadre de l'« Assessment for Learning » (Black et Wiliam, 1998, 2009 ; Wiliam, 2011), que l'on peut rapprocher dans le champ francophone, de l'évaluation formative, les feedbacks ont explicitement pour vocation de soutenir la régulation des apprentissages des apprenants. En effet, les feedbacks s'inscrivent alors dans une scénarisation pédagogique au cœur de laquelle les temps d'évaluation viennent surtout informer apprenants et enseignants sur le degré d'acquisition des apprentissages visés, en vue d'ajuster les stratégies et conditions de ces mêmes apprentissages (Thompson et Wiliam, 2007). Parmi les modèles existants de l'apprentissage autorégulé (Panadero, 2017), la contribution de Butler et Winne (1995) demeure particulièrement représentative du champ. Dans leur modèle, les auteurs mettent l'emphase sur le système cognitif de l'apprenant. Ils exposent un système au cœur duquel une tâche et son contexte de réalisation sont interprétés par l'apprenant, à travers ses croyances motivationnelles et ses connaissances par rapport au domaine

de l'activité, notamment. Cette interprétation amène l'apprenant à (re)configurer des buts d'accomplissement et les stratégies à appliquer pour les atteindre. Ces dernières permettent à l'apprenant de réaliser la tâche demandée, qui, une fois évaluée, devient une performance. À ce stade, les auteurs expliquent que l'apprenant engage, au cœur de son système cognitif, des actions de monitoring, c'est-à-dire de contrôle de la pertinence des buts définis, ou de l'efficacité des stratégies adoptées, par exemple. Ce faisant, l'apprenant génère un feedback interne, plus ou moins élaboré, et plus ou moins fiable selon qu'il ou elle mobilise des critères cohérents. En marge du système cognitif de l'apprenant, la performance déterminée par l'évaluation externe de la tâche donne lieu à un feedback externe, qui portant sur la tâche réalisée, permet à l'apprenant de confronter son feedback interne à celui de l'évaluateur, de façon à pouvoir ajuster ses croyances, ses connaissances, ses buts et/ou ses stratégies d'apprentissage. Si le modèle de Butler et Winne (1995) date déjà de trente années au moment où nous rédigeons le présent chapitre, il reste très pertinent, comme en attestent notamment les récents travaux de Nicol (2020), pour illustrer à la fois la complexité du système cognitif au travers duquel une tâche est interprétée par l'apprenant, mais également la fonction première de la régulation des apprentissages qui vise à réduire l'écart entre les buts fixés et la performance établie, ce dont atteste le feedback externe. Cette fonction a conduit une série de chercheurs à se questionner sur les conditions auxquelles les feedbacks pouvaient assurer au mieux cet ajustement de la part des étudiants.

III. Le paramétrage de l'information et ses limites

La littérature scientifique prodigue aux enseignants et formateurs une série de recommandations quant aux pratiques à privilégier, ou à éviter, lorsqu'il s'agit de fournir des feedbacks à des apprenants. Ainsi, Hattie et Timperley (2007) préconisent de couvrir des aspects qui dépassent le seul niveau de compréhension ou de performance atteint par un étudiant ou un élève, pour remettre cette information de base en perspective avec les objectifs visés (« Feed-up »). L'idée est alors de clarifier, voire d'objectiver, l'écart (« Feed-back ») qui, très probablement, aura été constaté par l'évaluateur. En outre, les auteurs enjoignent les praticiens à proposer des pistes d'amélioration (« Feed-forward ») sur la manière de combler cet écart, afin d'inciter l'apprenant à entreprendre des actions concrètes visant à progresser. Au-delà du spectre des informations à couvrir, Hattie et Timperley (2007) précisent que ces dernières peuvent porter sur la qualité

de la tâche réalisée (ex. : via la note, une appréciation), sur le processus ou la manière dont la tâche a été réalisée, sur la régulation qui a pu être opérée par l'apprenant préalablement à la tâche ou qu'il pourrait mettre en œuvre à la suite de celle-ci, et enfin, sur la personne de l'étudiant, c'est-à-dire sur ses qualités, ses lacunes ou encore sur les efforts consentis pour réaliser la tâche demandée. Pour compléter notre propos, précisons encore que les auteurs conseillent d'éviter autant que possible de centrer l'information sur le niveau personnel, pour tendre davantage vers le processus de réalisation de la tâche et la (ou les) piste (s) de régulation.

De telles recommandations ne sont pas les seules que l'on puisse trouver dans la littérature. Ainsi, les contributions de Nicol et Macfarlane-Dick (2006), de Shute (2008) ou encore de Brookhart (2008) sont également reconnues comme de précieuses balises auxquelles se fier dans la pratique du feedback. Les premiers ont proposé sept principes soutenant de bonnes pratiques de feedback. De sept, Nicol a fait passer ces principes à dix (Nicol, 2007), puis à douze dans leur version la plus récente (Nicol, 2009). Ils impliquent notamment le soutien motivationnel des apprenants ou encore la prise en compte par les enseignants eux-mêmes des informations délivrées, dans la régulation de leurs enseignements. Shute (2008) a établi un listing de pratiques à mettre en œuvre, et d'autres à proscrire, ainsi que des recommandations propres aux apprenants présentant des difficultés d'apprentissage, notamment. Enfin, Brookhart (2008, 2017) a organisé ses recommandations en les orientant tantôt vers le contenu des feedbacks (ex. la clarté du message, sa valence), et tantôt vers les stratégies (ex. : le timing, le mode de diffusion) propres à les communiquer aux apprenants.

Cela étant, ces ressources semblent se heurter depuis près de vingt ans à une série de limites, qui se concrétisent dans une sous-exploitation des feedbacks par les étudiants (Beaumont et al., 2011 ; Jonsson, 2013). Bien que ces aspects aient été étudiés de longue date par la communauté scientifique (Brookhart, 2008, 2017 ; Nicol, 2009 ; Nicol et Macfarlane-Dick, 2006) cette sous-exploitation renvoie inévitablement à la question des conditions de mise en œuvre des feedbacks (ex. : les feedbacks sont-ils suivis d'opportunités permettant aux apprenants d'en tenir compte ? Les informations fournies l'ont-elles été de manière suffisamment claire pour les apprenants ?) mais aussi à celle de la résonance que ces dernières peuvent provoquer sur les plans motivationnel et émotionnel, notamment. En effet, les expériences au cours desquelles les feedbacks sont perçus comme inutiles (Boud et Molloy, 2013 ; Carless, 2006), où ils plongent les étudiants dans des états émotionnels particulièrement inconfortables (Hill et al., 2021 ; Shields, 2015), où ils ne se sentent pas suffisamment compris ni même considérés, laissent des traces qui peuvent être persistantes en

termes de réponse de l'apprenant aux feedbacks (Lipnevich et al., 2016). Face à de telles difficultés, une autre question peut se poser, celle de la conceptualisation des feedbacks eux-mêmes ou, autrement dit, de se demander si l'information constitue bien un feedback où s'il peut s'agir d'autre chose.

En effet, d'après Sadler (2010), *« le problème fondamental réside moins dans la qualité des feedbacks que dans l'hypothèse selon laquelle le fait de simplement dire les choses, même de façon détaillée, constitue la voie la plus appropriée pour améliorer un apprentissage complexe. Apprendre en étant simplement informé est une stratégie douteuse, car les conditions pour que ces commentaires soient intégrés dans le travail de l'étudiant (en vue d'un travail futur) sont rarement remplies »* (traduit de Sadler, 2010, p. 539). Les propos de Sadler (2010), qui trouvent par ailleurs un écho dans le domaine de l'« Assessment for Learning » (Sambell et al., 2012), mettent à mal la dynamique unidirectionnelle qui caractérise la transmission d'informations d'une personne qui évalue la réalisation d'une tâche, vers la personne qui a réalisé cette même tâche. Ils soulignent la nécessité d'aller plus loin que cette seule transmission, afin de permettre aux apprenants de s'impliquer de manière plus directe dans la régulation de leurs apprentissages. C'est dans ce contexte que le concept de feedback a récemment évolué, pour s'ouvrir à une compréhension quelque peu différente de la transmission d'informations d'un émetteur vers un récepteur, sans doute devenue trop simple, et trop risquée. En effet, comme le souligne Romainville (voir aon chapitre dans le présent ouvrage), le feedback peut être différemment interprété selon que l'on considère sa dimension locutoire (ex. commentaire sur la qualité du travail réalisé ou de la performance évaluée) ou sa dimension perlocutoire (ex. ce que les étudiants comprennent du feedback, de ce qui doit en être fait, de ce sur quoi il porte).

IV. L'émergence de la « Student Feedback Literacy »

S'appuyant sur les travaux de Lea et Street (1998) sur les « Academic Literacies » et de ceux de Price et al. (2012) sur l'« Assessment Literacy », Sutton (2012) a développé le concept de « Feedback Literacy ». Celle-ci consiste en une série de connaissances et d'aptitudes spécifiques, inhérentes aux étudiants, leur permettant de décoder, d'interpréter et d'utiliser les feedbacks qui leur sont adressés. Quelques années plus tard, Carless et Boud (2018) ont repris ce concept et l'ont étendu aux principales

connaissances et aptitudes que les étudiants doivent développer pour jouer un rôle (pro)actif dans le traitement des feedbacks externes. Ce faisant, les feedbacks externes se repositionnent dans le processus de régulation des apprentissages comme une partie d'un processus plus complexe (O'Donovan et al., 2021), tel que l'avaient déjà présenté Butler et Winne (1995). Le modèle de Carless et Boud (2018) illustre depuis sa publication ce que d'aucuns ont appelé le « nouveau paradigme » du feedback (Carless, 2015 ; Dawson et al., 2019), par volonté de distinguer le feedback de la seule transmission d'informations, que ces auteurs décrivent comme emblématique d'un « ancien paradigme ». Cette conceptualisation alternative a suscité l'intérêt de nombreux chercheurs (Nieminen et Carless, 2023), qui y ont vu une réponse prometteuse aux limites associées aux pratiques de feedback telles qu'elles ont été relevées dans la littérature scientifique. Dans le cadre de ce nouveau paradigme, Carless et Boud (2018, p. 1315) proposent de définir le feedback comme *« […] un processus par lequel les étudiants font sens des informations qu'ils reçoivent d'une variété de sources et les utilisent pour améliorer leur travail ou leurs stratégies d'apprentissage »* (traduit de Carless et Boud, 2018, p. 1315).

Il est aisé de constater qu'à travers cette définition, les auteurs ont voulu mettre à distance les modèles les plus influents dans le domaine des feedbacks, qui plaçaient l'information en leur centre. Ce faisant, a contrario des modèles relevant de l'ancien paradigme, Carless et Boud (2018) ont souhaité marquer un déplacement de focale dans la définition du feedback, en la plaçant cette fois sur le rôle de l'apprenant dans l'interprétation et l'utilisation des feedbacks, qu'ils définissent désormais comme un processus, et non plus comme un produit.

Le modèle de la « Student Feedback Literacy » (Carless et Boud, 2018, p. 1319) postule la maîtrise, par l'apprenant, de quatre aptitudes :

- Apprécier les feedbacks :
    - Comprendre et reconnaître le rôle des feedbacks dans l'amélioration des apprentissages et le rôle actif qu'il ou elle doit jouer dans ce processus ;
    - Reconnaître les différentes formes et sources d'information des feedbacks ;
    - Utiliser les outils technologiques pour accéder aux feedbacks, les conserver et y revenir ultérieurement.

- Établir des jugements évaluatifs :
    - Déterminer la qualité de son travail/de sa performance (autoévaluation) et/ou de celui/celle de tiers ;
    - Participer de façon constructive à des dispositifs de feedbacks par les pairs ;
    - Affiner constamment ses capacités d'autoévaluation pour établir des jugements plus robustes.
- Gérer les affects :
    - Maintenir un équilibre émotionnel face aux critiques afin d'éviter d'adopter une attitude défensive à leur égard ;
    - Solliciter proactivement des suggestions de la part de ses pairs et/ou des enseignants, et entretenir un dialogue avec eux ;
    - Chercher à s'améliorer constamment à partir des feedbacks internes et externes.
- Engager une action :
    - Être conscient de la nécessité d'engager une action en réponse aux feedbacks ;
    - Réaliser des inférences à partir d'une variété de feedbacks pour chercher à progresser constamment ;
    - Développer un répertoire de stratégies pour agir sur base des feedbacks.

Il importe de remarquer que le modèle de Carless et Boud (2018) place ces différentes aptitudes à deux niveaux distincts. L'appréciation des feedbacks, l'établissement de jugements évaluatifs et la gestion des affects se situent à un premier niveau et sont associés par des liens d'influence mutuelle. Autrement dit, les auteurs mettent en évidence le lien qui existe entre la compréhension que les étudiants ont des feedbacks, la perception qu'ils en ont à travers le jugement évaluatif qu'ils posent dessus et les émotions suscitées par ces deux premières composantes. Par ailleurs, les auteurs pointent l'influence des émotions des étudiants sur ces deux autres composantes, indiquant qu'elles peuvent modifier les croyances des étudiants quant aux feedbacks (Adcroft, 2011 ; Dawson et al., 2019 ; Lizzio et Wilson, 2008 ; Orsmond et Merry, 2011), et la manière dont ils les utilisent pour apprécier la qualité de leurs productions ou performances. De cette manière, la combinaison des trois premières aptitudes augmente

ou diminue la probabilité que l'apprenant mobilise la quatrième composante du modèle, à savoir la mise en œuvre d'actions concrètes afin d'améliorer son travail et/ou les stratégies que celui-ci requiert.

D'après Winstone et ses collègues (2022, p.59), *« la Feedback Literacy permet aux étudiants de mieux utiliser les opportunités de feedback existant dans les programmes universitaires et les dispositifs d'évaluation à court terme, et les aide à développer leur jugement évaluatif et à s'engager plus efficacement dans l'environnement d'apprentissage/de travail à plus long terme »*. Il apparaît dès lors que l'enjeu de cette manière de concevoir les feedbacks et des aptitudes y afférentes est d'outiller les apprenants afin qu'ils soient en mesure de tirer le meilleur profit des informations qu'ils génèrent et/ou reçoivent dans le cadre de leurs activités d'apprentissage. Ce faisant, il s'agit de pallier des limites importantes (ex. : le manque d'engagement dans la recherche et l'interprétation de feedbacks externes) de pratiques considérées comme dépassées, ou à tout le moins inefficaces, dès lors qu'elles reposent exclusivement sur une démarche de transmission d'informations des enseignants vers les apprenants.

Cette conception alternative des feedbacks pourrait donner l'impression que l'efficacité des feedbacks repose uniquement sur les épaules des étudiants. Comme si leur « mission », désormais, était de se débrouiller pour acquérir les connaissances et les aptitudes favorisant le traitement et l'utilisation des feedbacks qui leur sont adressés, et la recherche proactive de feedbacks auprès de sources diverses. Il est évident que la situation est plus complexe qu'elle n'y paraît.

Il ne s'agit en rien de gommer les enseignants du processus de feedback, ni même de nier la valeur informative (Steelman et al., 2004) du message que les étudiants pourront traiter afin de réguler leurs apprentissages. Même en modifiant la perspective par laquelle nous pouvons considérer les feedbacks, la « Student Feedback Literacy » (Carless et Boud, 2018) ne peut ni ne doit occulter cette base fondamentale. C'est dans cette optique qu'a été proposé le modèle « Teacher Feedback Literacy » (Gravett et al., 2020 ; Winstone et Carless, 2020) qui se positionne en miroir du modèle dévolu aux apprenants. Il s'agit d'envisager le processus de feedback dans une logique de partenariat ou encore d'alliance pédagogique (Ajjawi et Regehr, 2019) entre étudiants et enseignants. Autrement dit, la responsabilité de l'efficacité des feedbacks est alors clairement partagée entre les protagonistes des situations évaluatives. Si la « Student Feedback Literacy » (Carless et Boud, 2018) propose d'établir la liste des composantes propres à soutenir le traitement des feedbacks du côté des apprenants, la « Teacher Feedback Literacy » (Carless et Winstone, 2020)

traite quant à elle des enseignants qui se devraient, eux aussi, de développer des aptitudes spécifiquement dévolues aux pratiques de feedback (Boud et Dawson, 2021). Ces auteurs définissent ainsi la « Teacher Feedback Literacy » comme « *[…] les connaissances, l'expertise et les dispositions nécessaires pour concevoir des processus de feedback permettant aux étudiants d'assimiler le feedback et de favoriser le développement de la Feedback Literacy des étudiants* » (traduit de Carless et Winstone, 2020, p.4). Cette définition fait apparaître au cœur de la « Teacher Feedback Literacy » la fonction de support des apprentissages des étudiants et du développement de leurs propres aptitudes opérer un traitement optimal des feedbacks, soit produire, rechercher, recevoir, interpréter et utiliser ceux-ci pour améliorer leur travail et/ou les stratégies qu'ils y déploient.

Boud et Dawson (2021) proposent les aptitudes suivantes comme principales composantes de la « Teacher Feedback Literacy » :

- Concevoir les feedbacks pour qu'ils soient utilisés : paramétrer l'information en tenant compte des contenus et stratégies possibles, de façon à rendre celle-ci exploitable.
- Se montrer sensible aux aspects relationnels : tenir compte de la relation pédagogique et de ses manifestations, notamment sur le plan affectif, afin de construire une relation de confiance entre l'enseignante et l'apprenant.
- Gérer les aspects pratiques : exploiter les ressources disponibles tout en considérant les contraintes inhérentes aux situations d'échanges entre l'enseignant et l'apprenant.

Ces quelques aptitudes doivent être mobilisées par les enseignants afin de permettre l'engagement des apprenants dans le traitement des feedbacks, de façon à pouvoir assumer davantage leur part de responsabilité dans le processus de régulation de leurs apprentissages. Ce faisant, cet engagement mobilise diverses dimensions qui attestent de la complexité des situations de feedback.

V. Le caractère multidimensionnel du traitement des feedbacks

Aujourd'hui, les recherches portant sur les feedbacks s'attèlent donc pleinement à comprendre le rôle de l'apprenant dans le processus que constitue le traitement des feedbacks, et à identifier les conditions dans

lesquelles ce rôle peut s'exprimer au mieux. En conséquence, il importe de se pencher sur l'apprenant, dans toute sa complexité, pour saisir plus finement le mécanisme par lequel il opère un traitement sur les informations qu'il génère lui-même, et sur celles qu'il reçoit de sources extérieures. Cela nécessite de s'intéresser à plusieurs dimensions qui, de façon concomitante, influencent le cours du traitement considéré.

La dimension cognitive est bien sûr engagée chez l'apprenant, dès lors qu'il effectue une série d'opérations mentales dès qu'une tâche lui est présentée, tel que nous avons pu le constater dans le modèle de l'apprentissage autorégulé de Butler et Winne (1995). En outre, la dimension émotionnelle fait à présent pleinement partie de l'équation que cherchent à résoudre les travaux dédiés aux feedbacks. Il apparaît dans diverses publications que celles-ci occupent un rôle pivot dans le traitement que les apprenants doivent réaliser. Elles engagent elles-mêmes une composante cognitive forte, à la fois dans le déclenchement et la réponse à des épisodes émotionnels fréquemment provoqués par les feedbacks dans les contextes éducatifs (Goetz et al., 2018 ; Hausman, 2024 ; Pekrun et al., 2006). Les émotions des apprenants dans les situations de feedbacks sont influencées par l'analyse que ceux-ci font de ladite situation, de son contexte et de ses conséquences réelles ou supposées. Parmi les variables qui entrent en ligne de compte, la relation aux tiers est importante. Qu'il s'agisse d'identifier des sources de jugement, de critiques additionnelles, ou à l'inverse de valorisation ou de soutien, les autres individus concernés de façon directe (ex. : l'enseignant ou les pairs le cas échéant) ou indirecte (ex. : les proches) par le feedback pèsent dans la « lecture » que l'apprenant fait de la situation. Ainsi, le degré de confiance que l'apprenant manifeste à l'égard de l'enseignant qui lui témoigne son soutien dans le cadre d'un feedback plus difficile à recevoir va influencer sa réaction, et notamment le fait d'y réagir de façon constructive ou de le rejeter, par exemple. De cette manière, la dimension sociale est, elle aussi, bien présente dans le traitement des feedbacks. Enfin, la dimension comportementale est évidemment intégrée au traitement des feedbacks dès lors qu'elle inclut ce que fait l'apprenant pour obtenir, donner du sens, et mettre en œuvre des feedbacks. La finalité, pointée par Carless et Boud (2018) dans leur modèle, est par ailleurs la mise en action de l'apprenant, sur base des informations qu'il ou elle aura pu mobiliser dans une optique d'autorégulation.

Comprendre les conditions d'efficacité des feedbacks nécessite donc une approche plus large du processus que constitue le traitement des informations générées et/ou reçues par les apprenants, et pas seulement une focalisation sur le message qui leur est adressé.

VI. Un regard critique sur l'approche contemporaine des feedbacks en éducation

Si le modèle de la « Student Feedback Literacy » de Carless et Boud (2018) incarne une nouvelle mouvance dans le champ des feedbacks en éducation, il convient de préciser qu'il présente malgré tout certaines limites. En effet, ce modèle est le fruit d'un travail conceptuel et s'appuie à ce titre sur un raisonnement mené à partir des résultats proposés par d'autres travaux de recherche dans le domaine. Et si ce modèle a déjà séduit de nombreux chercheurs, les articles empiriques qui permettent de valider ce modèle sont encore rares (Nieminen et Carless, 2023). Cependant, des outils ont été développés pour apprécier le développement de la « Feedback Literacy » chez les apprenants, en particulier dans le contexte de l'enseignement supérieur (Dawson et al., 2023 ; Dong et al., 2023 ; Zhan, 2022). Gageons que l'utilisation de ces outils permettra à l'avenir de mieux objectiver le rôle des apprenants et des aptitudes qu'ils développent dans le domaine des feedbacks, dans l'efficacité du traitement à opérer sur les informations relatives à leurs apprentissages.

Bien que cette approche récente du feedback à proprement parler a suscité un intérêt, et même un enthousiasme certain chez de nombreux chercheurs, elle a également entraîné certains d'entre eux à exprimer une forme de prudence quant à la manière de positionner le feedback en tant que processus (Lipnevich et Panadero, 2021 ; Panadero et Lipnevich, 2022). Ainsi, une compréhension du feedback comme étant l'information qui est générée en réponse à une action relative à un apprentissage, considérant dès lors que la prise en compte de celle-ci relève alors de la régulation ou de l'autorégulation de ce même apprentissage, est également largement défendue.

Au fil des travaux que nous avons menés dans ce domaine (Hausman et al., 2022, 2023a, 2023b), nous avons régulièrement cherché à clarifier notre positionnement vis-à-vis de la manière de conceptualiser les feedbacks. Sans y revenir en détail dans le présent chapitre, nous avons notamment souligné notre volonté de ne pas tomber dans le piège d'une lecture dichotomique de ces propositions, opposant les deux « écoles ». Nous pensons qu'il est préférable de chercher une voie médiane comme alternative à cette possible opposition. Cette troisième voie, que nous avons tenté d'adopter au long de ce chapitre, comprend comme processus le traitement que les étudiants appliquent aux feedbacks qu'ils génèrent

(internes) ou qu'ils reçoivent (externes). Le traitement représente le trait d'union entre l'information consécutive à la réalisation d'une tâche et l'engagement des apprenants dans sa régulation. Nous avons cherché, de cette manière, à minimiser le risque d'ambiguïté qui entoure ce concept, en utilisant plus fréquemment le terme de traitement des feedbacks (Lipnevich et Panadero, 2021) – voire de traitement constructif des feedbacks (Fong et al., 2018a, 2018b) – plutôt que d'évoquer les feedbacks seuls, dès lors que l'on s'intéresse à ce que les apprenants en font.

Afin de ne pas perdre de vue que le feedback demeure un moyen de faire évoluer les apprenants, et non une fin en soi, nous avons insisté sur l'intégration du traitement des feedbacks dans un système plus large qui est celui de la régulation des apprentissages. Comme nous l'écrivions précédemment (Hausman, 2024, p.254) : *« Nous sommes en effet convaincus que c'est en insistant sur ce type de dénominations que la communauté scientifique parviendra à dépasser un débat (Nieminen et Carless, 2023 ; Panadero et Lipnevich, 2022 ; Winstone et al., 2022) qui, d'un côté comme de l'autre, restreint le champ des possibles et limite la portée de ce concept clé dans le domaine de l'éducation »*.

VII. Bibliographie

Adcroft, A. (2011). The mythology of feedback. *Higher Education Research & Development*, *30*(4), 405-419.

Ajjawi, R. et Regehr, G. (2019). When I say… feedback. *Medical Education*, *53*(7), 652-654.

Beaumont, C., O'Doherty, M. et Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education*, *36*(6), 671-687.

Boud, D. et Dawson, P. (2021). What feedback literate teachers do: an empirically-derived competency framework. *Assessment & Evaluation in Higher Education*, *48*(2), 158-171.

Boud, D. et Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698-712.

Black, P. et Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Black, P. et Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*, 5-31.

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Ascd.

Brookhart, S. M. (2017). *How to give effective feedback to your students*. Ascd.

Butler, D. L. et Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Carless, D. (2006). Differing Perceptions in the Feedback Process. *Studies in Higher Education, 31*(2), 219-233.

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. (1e éd.). Routledge.

Carless, D. et Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325.

Carless, D. et Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, *28*(1), 150–163.

Dawson, P., Carless, D. et Lee, P. P. W. (2021). Authentic feedback: supporting learners to engage in disciplinary feedback practices. *Assessment & Evaluation in Higher Education*, *46*(2), 286-296.

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D. et Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, *44*(1), 25-36.

Dawson, P., Yan, Z., Lipnevich, A., Tai, J., Boud, D. et Mahoney, P. (2023). Measuring what learners do in feedback: the feedback literacy behaviour scale. *Assessment & Evaluation in Higher Education*, 1-15.

Detroz, P., Malay, L. et Crahay, V. (2020). Une démarche structurée pour définir quelques conseils limitant l'impact de la pandémie sur l'évaluation de nos étudiants. *Évaluer : Journal International de Recherche en Education et Formation, 1*, 97-110.

Dong, Z., Gao, Y. et Schunn, C. D. (2023). Assessing students' peer feedback literacy in writing: Scale development and validation. *Assessment & Evaluation in Higher Education*, *48*(8), 1103-1118.

Fong, C. J., Schallert, D. L., Williams, K. M., Williamson, Z. H., Warner, J. R., Lin, S. et Kim, Y. W. (2018a). When feedback signals failure but offers hope for improvement: A process model of constructive criticism. *Thinking Skills and Creativity, 30*, 42-53.

Fong, C. J., Williams, K. M., Williamson, Z. H., Lin, S., Kim, Y. W. et Schallert, D. L. (2018b). "Inside out": Appraisals for achievement emotions from constructive, positive, and negative feedback on writing. *Motivation and Emotion, 42*, 236-257.

Goetz, T., Lipnevich, A., Krannich, M. et Gogol, K. (2018). Performance Feedback and Emotions. Dans A. Lipnevich et J. Smith (dir.), *The Cambridge Handbook of Instructional Feedback* (p. 554-574). Cambridge University Press.

Gravett, K., Kinchin, I. M., Winstone, N. E., Balloo, K., Heron, M., Hosein, A., Lygo-Baker, S. et Medland, E. (2019). The development of academics' feedback literacy: experiences of learning from critical feedback via scholarly peer review. *Assessment &Evaluation in Higher Education*, *45*(5), 651–665.

Hausman, M. (2024). De la cognition aux émotions : étude compréhensive du traitement des feedbacks par les étudiants de l'enseignement supérieur. [thèse de doctorat non publiée], Université de Liège.

Hausman, M., Dancot, J., Pétré, B., Guillaume, M. et Detroz, P. (2022). 'I don't know if people realize the impact of their words': how does feedback during internship impact nursing student learning? *Assessment & Evaluation in Higher Education, 48*(6), 777-789.

Hausman, M., Detroz, P., Pétré, B., Guillaume, M. et Dancot, J. (2023a). Feedback processing and emotion regulation in nursing students during internship. *Learning and Instruction, 87*, 101804.

Hausman, M., Leduc, L., Malay, L., Delvaux, S. et Detroz, P. (2023b). Analysis of the Emotions and Emotional Skills of University Students in Processing Distance Formative Feedback. *Mesure et Évaluation en Éducation*, *46*, 22-54.

Hattie, J. et Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112.

Hill, J., Berlin, K., Choate, J., Cravens-Brown, L., McKendrick-Calder, L., et Smith, S. (2021). Exploring the emotional responses of undergraduate students to assessment feedback: Implications for instructors. *Teaching and Learning Inquiry: The ISSOTL Journal, 9*(1), 294-316.

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education, 14*(1), 63-76.

Kluger, A. N, et DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Kulhavy, R. W. et Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279-308.

Lea, M. R. et Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, *23*(2), 157-172.

Lipnevich, A. A., Berg, D. A. et Smith, J. K. (2016). Toward a model of student response to feedback. Dans G. T. L. Brown et L. R. Harris, (dir.), *Handbook of human and social conditions in assessment* (p. 169-185). Routledge.

Lipnevich, A. A. et Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*, 720195.

Lizzio, A. et Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, *33*(3), 263-275.

Nicol, D. (2007, 29 au 31 mai). *Principles of good assessment and feedback: Theory and practice* [communication orale]. REAP International Online Conference on Assessment Design for Learner Responsibility.

Nicol, D. (2009). Quality enhancement themes: The first year experience. *Transforming assessment and feedback: enhancing integration and empowerment in the first year.* The Quality Assurance Agency for Higher Education.

Nicol, D. (2020). The power of internal feedback: exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*(5), 756-778.

Nicol, D. J. et Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218.

Nieminen, J. H. et Carless, D. (2023). Feedback literacy: A critical review of an emerging concept. *Higher Education, 85*(6), 1381-1400.

O'Donovan, B., den Outer, B., Price, M. et Lloyd, A. (2021). What makes good feedback good? *Studies in Higher Education*, *46*(2), 318-329.

Orsmond, P. et Merry, S. (2011). Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, *36*(2), 125-136.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422.

Panadero, E. et Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*, 315-341.

Pitt, E. et Norton, L. (2017). 'Now that's the feedback I want! 'Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, *42*(4), 499-516.

Price, M., Handley, K. et Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, *36*(8), 879-896.

Price, M., C. Rust, B. O'Donovan, K. Handley et R. Bryant. (2012) Assessment Literacy: The Foundation for Improving Student Learning. OCSLD.

Sadler, D. R. (2010) Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535-550.

Sambell, K., McDowell, L. et Montgomery, C. (2012). *Assessment for Learning in Higher Education*. Routledge.

Shields, S. (2015). 'My work is bleeding': exploring students' emotional responses to first-year assignment feedback. *Teaching in Higher Education, 20*(6), 614-624.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research, 78*(1), 153-189.

Steelman, L. A., Levy, P. E. et Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement*, *64*(1), 165-184.

Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International*, *49*(1), 31-40.

Thompson, M. et Wiliam, D. (2007). Tight but loose: A conceptual framework for scaling up school reforms. Dans *annual meeting of the American Educational Research Association. Chicago, IL*.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3-14.

Winstone, N. E., Balloo, K. et Carless, D. (2022). Discipline-specific feedback literacies: A framework for curriculum design. *Higher Education*, *83*(1), 57-77.

Winstone, N. et Carless, D. 2020. *Designing Effective Feedback Processes in Higher Education: A Learning-Focused Approach.* (1e éd.). Routledge.

Zhan, Y. (2022). Developing and validating a student feedback literacy scale. *Assessment & Evaluation in Higher Education*, *47*(7), 1087-1100.

Chapitre 19. Culture, pratiques et évolution de l'évaluation en éducation et formation

Quarante années de transformation et de recherche au Maroc

Ghizlane Chemsi, Naceur Achtaich, Mohamed Radid et Mohammed Talbi

I. Introduction

L'évaluation en éducation et formation constitue aujourd'hui un enjeu majeur pour les systèmes éducatifs à travers le monde. Elle est perçue non seulement comme un outil de mesure des acquis des apprenants, mais aussi comme un levier essentiel pour améliorer la qualité des enseignements, piloter les politiques éducatives et garantir la régulation, la transparence et la responsabilité des institutions. Elle s'inscrit désormais dans une logique de développement, où l'objectif n'est plus seulement de contrôler, mais d'accompagner et de transformer les apprentissages, les pratiques ou encore les institutions.

Au Maroc, cette dynamique s'inscrit dans un contexte de réformes profondes et de transformations ambitieuses visant à répondre aux défis du développement socio-économique et à promouvoir une éducation inclusive, équitable et performante. L'évaluation en éducation a connu une évolution significative sous l'effet des réformes institutionnelles successives, des exigences de gouvernance et de l'essor de la recherche

académique, mais aussi de l'émergence de nouveaux environnements numériques d'apprentissage (OCDE, 2024 ; Akrim et al., 2010).

L'histoire de l'évaluation dans le système éducatif marocain est relativement récente et a connu plusieurs phases d'évolution. Depuis la fin des années 1970, un intérêt croissant s'est manifesté pour la mesure des performances scolaires et l'analyse de l'efficacité des programmes de formation. Cependant, ces premières initiatives, souvent ponctuelles et isolées, n'ont pas permis d'instaurer une culture d'évaluation durable et systématique. Ce n'est qu'à partir des années 1980, avec la mise en place de dispositifs institutionnels spécifiques, que l'évaluation a commencé à s'inscrire dans une perspective plus structurée et stratégique.

Cette évolution a été consolidée par les initiatives portées par les instances éducatives et soutenue par l'appui d'agences internationales. L'intégration de l'évaluation comme outil de pilotage et d'amélioration continue de la qualité a été marquée par plusieurs jalons majeurs : la création du Centre National d'Évaluation des Programmes en 1998, l'adoption de la Charte Nationale d'Éducation et de Formation en 1999, la Loi-Cadre 51-17 et la Vision stratégique 2015–2030. Ces réformes traduisent la volonté des autorités marocaines de professionnaliser durablement cette fonction.

Par ailleurs, les mutations pédagogiques ont placé l'évaluation au cœur des transformations des pratiques éducatives. L'introduction de l'approche par compétences (APC), la diversification des fonctions de l'évaluation (diagnostique, formative, sommative, certificative) et des modalités (critériée, par les pairs, numérique, présentielle/hybride), ainsi que l'évolution des pratiques enseignantes ont contribué à repositionner les dispositifs évaluatifs comme outils d'accompagnement des apprentissages (El Gharbi, 2019 ; Hajji, 2016). Ces évolutions exigent un renforcement des compétences professionnelles, mais aussi une reconfiguration des représentations : il ne s'agit plus d'évaluer pour sanctionner, mais pour faire progresser.

Dans cette dynamique, la contribution d'acteurs scientifiques et institutionnels a été déterminante. Les apports de l'ADMEE-Europe et de l'Observatoire de Recherche en Didactique et Pédagogie Universitaire (ORDIPU) ont permis de structurer la réflexion, de diffuser des cadres théoriques, et de favoriser les échanges de pratiques à travers des colloques, séminaires, écoles d'été et projets de recherche collaboratifs.

Ce chapitre se propose d'analyser la culture, les pratiques et l'évolution de l'évaluation en éducation et formation au Maroc, à partir d'une revue de littérature mobilisant des données issues de la recherche scientifique, des

textes institutionnels et des retours d'expériences. Une attention particulière est accordée aux travaux menés par l'Observatoire de Recherche en Didactique et Pédagogie Universitaire (ORDIPU) et par la section marocaine de l'ADMEE-Europe. Ce corpus est complété par d'autres recherches ayant contribué à consolider ce champ à l'échelle nationale et internationale. L'analyse vise à offrir une synthèse critique des avancées réalisées, à identifier les défis persistants, et à formuler des pistes pour renforcer la fonction évaluative dans un contexte de réforme continue. En croisant les regards de chercheurs, de praticiens et de décideurs, ce travail entend nourrir une réflexion collective sur un enjeu stratégique pour l'avenir du système éducatif marocain.

II. La culture évaluative dans le système éducatif marocain : réformes, acteurs et enjeux

La culture de l'évaluation en éducation au Maroc est en constante évolution, marquée par des avancées importantes mais aussi par des défis structurels et conceptuels persistants. L'évolution de la culture évaluative au Maroc témoigne d'un passage progressif d'une logique de contrôle normatif à une conception plus formative et stratégique de l'évaluation. Ce mouvement s'est construit au rythme des grandes réformes éducatives et de l'action de nouveaux acteurs institutionnels. Toutefois, cette dynamique reste confrontée à des défis conceptuels et organisationnels persistants. C'est à l'articulation entre ces réformes, les acteurs impliqués et les enjeux qui en découlent que s'intéressent les sections suivantes.

Durant la période de centralisation et de certification (1980–1999), l'évaluation était normée, descendante et orientée vers la sélection scolaire. Elle se matérialisait par des examens standardisés sans réelle fonction formative (Bennani, 1986).

Par la suite, les réformes fondatrices — la Charte nationale d'éducation et de formation (1999) et la Vision stratégique 2015–2030 — ont introduit l'approche par compétences (APC), la logique de pilotage par les résultats, et une reconnaissance de l'évaluation comme levier d'amélioration continue. L'Observatoire de Recherche en Didactique et Pédagogie Universitaire (ORDIPU), créé à l'Université Hassan II de Casablanca en 2002, a fortement contribué à cette dynamique en produisant des travaux sur les effets de l'APC et l'intégration de l'évaluation formative.

Dans cette continuité, le Maroc a également développé des dispositifs d'autoévaluation et d'évaluation externe des établissements scolaires, notamment à travers le Projet d'Établissement Intégré (PEI). Ces démarches visent à responsabiliser les écoles et à les inscrire dans un processus d'amélioration continue. Toutefois, les capacités limitées en évaluation, le manque de formation des acteurs et l'absence de référentiels clairs freinent encore le développement de ces pratiques (Maghnouj, 2018 et OCDE, 2024).

Parallèlement, le Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique (CSEFRS), via l'Instance Nationale d'Évaluation (INE), a professionnalisé depuis 2015 l'évaluation systémique, en produisant des rapports détaillés. Le Programme National d'Évaluation des Acquis des élèves (PNEA), notamment, permet d'analyser l'impact des orientations éducatives sur les performances scolaires. Plusieurs chercheurs affiliés à l'ADMEE-Europe (section Maroc) y ont activement contribué.

Concernant l'enseignement supérieur, le Maroc a longtemps souffert d'une faible capacité d'évaluation et de suivi de l'insertion professionnelle des lauréats, en raison d'un déficit d'outils de gouvernance et d'une absence de cadre stratégique global. Ce n'est qu'en 2014 qu'a été créée l'Agence Nationale d'Évaluation et d'Assurance Qualité de l'Enseignement Supérieur (ANEAQ), en vertu de la loi n° 80-12 du 31 juillet 2014, promulguée par Dahir. Cette agence a pour mission d'évaluer le système d'enseignement supérieur et de recherche scientifique afin de garantir sa qualité, pour le compte de l'État. Autonome et indépendante, elle dispose d'un dispositif technique d'évaluation complet. L'ANEAQ assure l'harmonisation des référentiels de formation et met en place une démarche globale d'évaluation interne et externe, conforme aux standards internationaux, au service du ministère, du CSEFRS et des universités publiques et privées.

Une première expérience d'autoévaluation des universités a eu lieu en 2020. Seuls 55 établissements issus de 7 universités y ont participé. La formation à l'évaluation institutionnelle reste également limitée, puisque seules 139 personnes relevant de 5 universités ont été formées.

Au niveau de l'enseignement fondamental, une réforme a été lancée en 2024 à travers le programme des « écoles pionnières », qui a concerné environ un tiers des écoles primaires et plus d'un million d'élèves. Ce programme a introduit des évaluations diagnostiques rigoureuses en début d'année scolaire (2024–2025), accompagnées de dispositifs de remédiation ciblés pour renforcer les compétences de base en arabe, en

mathématiques et en français. Les évaluations indépendantes menées dans ce cadre ont produit des données probantes, révélant que 66% des élèves n'avaient pas atteint le niveau attendu dans ces disciplines fondamentales.

Malgré ces avancées, la culture de l'évaluation reste encore largement dominée par une approche bureaucratique et sommative, centrée sur la notation, la sélection et le contrôle administratif. Cela limite l'autonomie des enseignants et freine l'innovation pédagogique. La compréhension de l'évaluation est souvent confuse, mêlant évaluation, mesure et notation, ce qui traduit l'absence d'une culture évaluative partagée, liée à des imprécisions conceptuelles et à des pratiques peu diversifiées.

Cette situation est aggravée par la centralisation du système qui entrave l'implication active des élèves dans leur apprentissage et ne favorise pas l'émergence d'une véritable ingénierie de l'évaluation. La culture dominante reste marquée par une logique de contrôle sommatif, au détriment d'une évaluation formative au service des apprentissages (Hajji et Hamdani, 2017). En outre, les résultats des évaluations sont encore insuffisamment exploités pour orienter les décisions éducatives ou améliorer les pratiques pédagogiques. La faiblesse de la communication entre évaluateurs, enseignants, élèves, parents et institutions compromet l'impact positif potentiel d'une évaluation intégrée à la dynamique pédagogique.

Dans cette perspective, il devient nécessaire de :

- Renforcer la formation et l'autonomie des enseignants afin qu'ils puissent mettre en œuvre des évaluations formatives, alignées avec les objectifs du curriculum, et mieux exploiter les résultats pour améliorer les apprentissages ;
- Développer une culture de l'évaluation formative, centrée sur l'implication des élèves et dépassant la simple logique de contrôle (Hamdani, 2022) ;
- Améliorer la communication et l'exploitation des données d'évaluation pour éclairer les décisions stratégiques en éducation, mobiliser les acteurs éducatifs et renforcer la confiance dans le système (Maghnouj, 2018 ; Anajar et al., 2015) ;
- Poursuivre la réforme des écoles pionnières, en l'adaptant aux réalités du système marocain et en intégrant les enjeux de formation initiale, d'autonomie des établissements et de pilotage global.

Ainsi au Maroc, la culture évaluative en éducation, aux plans pédagogique, institutionnel et stratégique, connaît une évolution encourageante vers davantage de rigueur, d'autonomie et d'orientation formative. Pour qu'elle puisse pleinement jouer son rôle dans l'amélioration de la qualité des enseignements et des apprentissages, il reste toutefois nécessaire de lever un certain nombre de freins d'ordre conceptuel, institutionnel et pratique.

III. Pratiques et dispositifs d'évaluation dans les établissements au Maroc

Les recherches sur les pratiques évaluatives réelles au Maroc pouvaient se compter sur les doigts d'une main. Les instances où ces recherches émergeaient restaient limitées et dépendaient parfois d'efforts personnels pour aboutir (Zoubaidi, 2014). Nous rappelons, dans ce sens, plusieurs rencontres organisées ces dernières décennies et ayant eu pour thème l'évaluation : d'abord le colloque organisé par l'ORDIPU, en collaboration avec l'ADMEE-Europe, à l'automne 2007, intitulé « Place de l'évaluation dans la stratégie de développement du système éducatif : bilan préliminaire, évaluation et perspective de la réforme LMD, volet L ». Les préoccupations de ce colloque concernaient l'évaluation des apprentissages — avec une attention particulière portée aux

compétences — l'évaluation des enseignements par les étudiants et la prise en compte de leurs points de vue, ainsi que l'évaluation des dispositifs d'enseignement et de formation au sein du système LMD.

Au printemps 2008, le Conseil supérieur de l'éducation, de la formation et de la recherche scientifique (CSEFRS), via l'Instance nationale d'évaluation (INE), a organisé une rencontre intitulée « L'évaluation des systèmes d'éducation et de formation dans tous ses états ». L'objectif de ce colloque était d'imprégner l'ensemble des acteurs du secteur d'une culture de l'évaluation et d'offrir un espace de réflexion sur des pratiques pertinentes, adaptées aux particularités du contexte marocain. La rencontre s'est articulée autour des axes suivants : les pratiques de l'évaluation en éducation et en formation ; la culture, les acteurs, les méthodologies et les finalités de l'évaluation ; le métier d'évaluateur (profil et exigences) ; et l'évaluation de la réussite scolaire.

L'ADMEE-Europe et sa section Maroc ont organisé plusieurs colloques, en collaboration avec l'ORDIPU. Au printemps 2009, un colloque intitulé « Nouvelles formes et nouveaux modèles d'évaluation » a notamment eu lieu. Le principal axe de débat a porté sur la contractualisation : il s'agissait de clarifier les responsabilités et d'engager les acteurs, tout en encourageant les enseignants à développer la recherche en évaluation à l'université, notamment dans le cadre des procédures d'accréditation. Le colloque de l'ADMEE-Europe tenu à Marrakech en 2014, « Cultures et politiques de l'évaluation en éducation et en formation », avait pour objet de questionner les pratiques et dispositifs d'évaluation à travers la diversité des contextes et des politiques. Il a porté à la fois sur l'évaluation des apprentissages, des enseignements et des dispositifs, sur les méthodologies et les outils employés, ainsi que sur les contextes socioculturels dans lesquels l'évaluation s'est inscrite. Le colloque de l'ADMEE-Europe organisé en 2020 à Casablanca a traité la thématique « Dispositifs et méthodologies émergentes en évaluation ».

L'intérêt s'est principalement porté sur les innovations afférentes aux dispositifs, aux pratiques, aux méthodologies et aux outils d'évaluation, sur l'intégration de l'évaluation dans le développement professionnel des enseignants, sur la promotion d'une évaluation au service de la qualité de l'éducation et de la formation, ainsi que sur les enjeux liés aux innovations technologiques.

Lors de ce colloque, les travaux présentés par les membres de la section marocaine de l'ADMEE-Europe montrent la résistance des enseignants à l'adoption de pratiques évaluatives formatives, et ce, en raison du manque

de formation et d'un climat scolaire peu propice à la régulation (Hajji et Hamdani, 2017).

Des expériences pilotes conduites par l'ORDIPU sur l'évaluation critériée, les portfolios d'apprentissage ou les tâches complexes ont montré des effets positifs sur l'engagement des étudiants. Ces expériences, qui placent la motivation et l'engagement au cœur des pratiques évaluatives, valorisent également des formats tels que les jeux sérieux, les feedbacks interactifs ou encore les évaluations individualisées (El Gharbi, 2019 ; Ouchaouka, 2019).

D'autres travaux ont établi que l'évaluation des enseignements par les étudiants constitue un levier pour la prise de décision pédagogique, à la fois par la conception et la validation d'outils et par la mise en évidence d'effets sur l'engagement (Chemsi et al., 2016 ; Lajane et al., 2021). Par ailleurs, l'intégration des technologies numériques, de l'IA et des learning analytics contribue à objectiver l'évaluation et à éclairer la prise de décision fondée sur des données (Housni et al., 2019). Dans certains dispositifs, ces outils soutiennent aussi la personnalisation des parcours et des rétroactions (Lajane et al., 2021 ; El Machtani El Idrissi et al., 2022 ; Naciri et al., 2023).

L'évaluation des dispositifs de formation des enseignants a fait également l'objet d'enquêtes menées par des chercheurs marocains qui soulignent l'écart entre les référentiels d'évaluation et les pratiques réelles (Essaoudi, 2012 et Khyati, 2013). Des travaux du Groupe d'évaluation de la section marocaine de l'ADMEE-Europe, en partenariat avec des Académies Régionales de l'Education et de La Formation (AREF) et des universités, ont proposé des grilles d'évaluation institutionnelle des établissements scolaires, fondées sur des indicateurs contextuels, pédagogiques et organisationnels.

Au niveau institutionnel, quelques autoévaluations ont eu lieu, principalement au niveau de l'enseignement supérieur. Nous citons deux exemples d'autoévaluation auxquels l'équipe de l'ORDIPU a apporté une contribution déterminante : d'abord, un exemple conduit dans le cadre d'un projet européen ; ensuite, un exemple réalisé sous l'égide de l'Agence nationale d'évaluation et d'assurance qualité (ANEAQ). À titre d'illustration, la Faculté des sciences Ben M'Sik (FSBM) de l'Université Hassan II de Casablanca (UH2C) a lancé, en 2015, un processus d'autoévaluation dans le cadre du projet européen TEMPUS (RECET) – Renforcement des compétences en évaluation institutionnelle, financé par l'Union européenne. Ce travail a permis d'identifier les atouts et les aspects à évaluer, en tenant compte des conditions internes et externes de

développement, et en distinguant points forts, axes d'amélioration, opportunités et points de vigilance. Dans la continuité, en 2020, la FSBM s'est engagée dans une nouvelle autoévaluation au titre d'un projet pilote de l'UH2C, en s'appuyant cette fois sur le référentiel national d'évaluation et d'assurance qualité de l'enseignement supérieur, conformément au décret n° 2-19-16 du 23 juillet 2019 pris pour l'application de l'article 5 de la loi n° 80-12 relative à l'ANEAQ.

Conformément au référentiel qualité de l'ANEAQ, l'auto-évaluation suit quatre niveaux hiérarchiques, du général au particulier : Domaines évalués, grandes catégories (gouvernance et management, fonctions supports, formation, recherche scientifique, accompagnement des étudiants et vie estudiantine, services à la société) ; Champs/dimensions, déclinaisons opérationnelles de chaque domaine précisant le périmètre de l'analyse ; Références, attendus structurants propres à chaque champ ; Critères, règles d'appréciation des références fondées sur des indicateurs et des preuves (questionnements, données, pièces justificatives).

Le rapport d'auto-évaluation établi propose, éventuellement, un suivi et un accompagnement de l'établissement dans ses plans d'actions. Lors de cette évaluation, on a constaté :

- La conformité de l'institution à une grande partie des exigences du référentiel, ce qui se traduit par un bon déroulement de l'ensemble des processus et permet l'atteinte d'une majorité des objectifs de la politique mise en place ;
- La volonté de la direction d'assurer un bon déroulement de cette évaluation et de maintenir davantage une politique basée sur l'amélioration continue ;
- La transparence et la disponibilité de l'ensemble des audités ainsi que leur maîtrise des différents processus ;
- La bonne cohérence entre les objectifs et le plan d'action mis en place par la direction ;
- L'envergure des investissements en termes de locaux et moyens pour pallier les problèmes constatés lors de la dernière évaluation en 2016 ;
- L'engagement de l'institution pour s'intégrer dans son environnement comme acteur potentiel de développement ;
- La volonté des différents acteurs pour participer au rayonnement de l'institution à l'échelle régionale, nationale et internationale.

Toutefois, des anomalies par rapport au référentiel (ANEAQ), qui ne mettent pas en cause le système de l'ensemble des processus, ont été constatées. Elles sont principalement liées à :

- L'absence d'une régularité en termes de maîtrise documentaire des procédures et d'enregistrements ;
- La maîtrise des canaux de communication qui permettront de meilleurs résultats en termes de diffusion de l'information ;
- La mise en place d'indicateurs efficaces et l'exploitation des données relatives à différentes enquêtes.

La consolidation de l'ensemble de points positifs et la mise en place des actions pour répondre à l'ensemble des écarts constatés par l'équipe d'évaluation ont permis à l'institution de se conformer à l'ensemble des exigences du référentiel et de maintenir à un haut niveau sa contribution au système éducatif national.

IV. La recherche scientifique en évaluation au Maroc : structuration, dynamiques et perspectives

La recherche en évaluation au Maroc a émergé lentement dans les années 1980 et 1990, essentiellement à travers des travaux de mémoire et de doctorat, sans structuration institutionnelle forte. Avec les réformes éducatives et la Charte nationale de l'éducation (1999), une dynamique s'est enclenchée autour de l'analyse critique des systèmes d'évaluation et de leurs impacts.

Dans cette logique de consolidation académique, la structuration de la recherche autour de pôles et réseaux spécialisés, la création de l'Observatoire de Recherche en Didactique et Pédagogie Universitaire (ORDIPU) en 2002 et de la section marocaine de l'ADMEE-Europe en 2009, ont favorisé l'émergence d'une communauté scientifique engagée. Les colloques de Marrakech (2014) et de Casablanca (2017, 2020) ont permis de consolider des axes de recherche, notamment sur l'éthique, les politiques évaluatives, la transformation pédagogique et l'évaluation numérique.

Malgré ces avancées, les recherches scientifiques menées au Maroc dans le domaine de l'évaluation éducative demeurent souvent fragmentées, ponctuelles et peu articulées aux réformes en cours. Elles souffrent d'un

déficit d'institutionnalisation, d'un manque d'archives longitudinales et d'une faible appropriation par les décideurs. D'où l'importance d'un travail de synthèse critique permettant de mieux cerner les avancées, les limites et les perspectives de la recherche en évaluation dans le contexte marocain.

Dans cette perspective et issue des réflexions menées par l'ORDIPU, la section marocaine de l'ADMEE-Europe s'est donné pour mission de produire des connaissances scientifiques contextualisées sur les dispositifs d'enseignement, de formation et d'évaluation, notamment dans l'enseignement supérieur. À travers une série de travaux empiriques conduits dans des contextes pédagogiques variés, les chercheurs de l'ORDIPU interrogent les dynamiques contemporaines de l'évaluation en éducation. Ces études, mobilisant des démarches qualitatives, quantitatives et mixtes, conçoivent l'évaluation comme un levier d'apprentissage, d'engagement et de transformation pédagogique. Elles déplacent ainsi la fonction évaluative de sa logique certificative vers une approche formative, participative et régulatrice, alignée sur les exigences d'équité cognitive.

Les travaux de l'ORDIPU révèlent une orientation méthodologique spécifique, centrée sur l'évaluation des apprentissages, des compétences, des dispositifs de formation et des orientations stratégiques du système éducatif. Parmi ces travaux, plusieurs études ont exploré les déterminants motivationnels et cognitifs de l'engagement des apprenants en lien avec les pratiques évaluatives (Naciri et al., 2023 ; Laaziz et al., 2024 ; Chakir, 2022). D'autres recherches, notamment psychométriques (Essaoudi et al., 2012), ont permis de développer et valider un outil d'évaluation de la qualité des établissements scolaires.

Au-delà des approches méthodologiques classiques, El Bousaadani (2010) a analysé les biais psychosociologiques de l'évaluation chez les acteurs impliqués dans la mise en œuvre des réformes. Les jugements sont influencés par plusieurs facteurs — sexe, niveau d'études, expérience professionnelle, responsabilité administrative, qualité des relations interpersonnelles et satisfaction perçue — les biais les plus marqués touchant les dimensions motivationnelles. El Bousaadani et al. (2010) montrent que ces facteurs peuvent induire des effets de sévérité/complaisance, d'affinité ou de halo, d'où la nécessité de former aux biais et de sécuriser les procédures (grilles critériées, triangulation des sources, anonymisation partielle, rétroactions).

De manière complémentaire, une étude de cas menée au Centre Régional des Métiers de l'Éducation et de la Formation (CRMEF) de Casablanca,

conduite par Terhzaz et al. (2015), a porté sur l'évaluation des enseignants par les étudiants (EEE) en sciences physiques. Cette recherche visait à identifier les effets de l'enseignement sur les élèves-professeurs, à diagnostiquer les éventuelles faiblesses du dispositif de formation et à instaurer une dynamique pédagogique fondée sur le dialogue et le partage.

Parallèlement, les mutations technologiques ont conduit les chercheurs à intégrer de nouvelles dimensions à leurs travaux. L'ORDIPU s'inscrit dans cette reconfiguration des pratiques évaluatives à travers les technologies numériques, les learning analytics et l'intelligence artificielle (Housni et al., 2019). Ces travaux visent à explorer dans quelle mesure ces outils peuvent enrichir l'évaluation formative, individualiser les parcours et promouvoir une pédagogie équitable et efficace. Des approches mixtes sont mobilisées, alliant analyses quantitatives (modélisation, trace-based analytics, prédiction) et qualitatives (expériences utilisateurs, feedbacks, engagement), appliquées à des contextes variés : formation infirmière, MOOC, SPOC, LMS et environnements immersifs (Ayat et al., 2024 ; Tilioui et al., 2025).

Les résultats de ces recherches soulignent l'importance d'intégrer l'évaluation au processus d'apprentissage (Chemsi et al., 2020 ; Ouchaouka et al., 2019), de l'ancrer sur des données exploitables en temps réel et de l'orienter vers l'adaptation pédagogique centrée sur l'engagement de l'apprenant (El Machtani et al., 2022 ; Moundy et al., 2021). Cette approche contribue à faire évoluer l'évaluation vers des pratiques hybrides, intelligentes et centrées sur l'action (Mrisse et al., 2023 ; Naciri et al., 2023). Elle illustre le potentiel transformateur des technologies pour concevoir des dispositifs d'évaluation adaptés aux environnements numériques d'apprentissage (Elimadi et al., 2024 ; Laaziz et al., 2024 ; Naciri et al., 2023), renforçant la qualité, l'équité et l'efficacité pédagogiques.

Les recherches de Chakir (2022) et Lajane et al. (2023) montrent que cette approche favorise la concentration, la motivation, la persévérance, l'autonomie et la réflexivité des apprenants, tout en illustrant les apports du mobile learning dans l'évaluation formative, notamment par le développement d'applications comme NursE-Quiz. Les travaux de Dari et al. (2024) abondent dans ce sens en soulignant l'importance de repenser l'ergonomie et la structuration des plateformes numériques pour une meilleure expérience d'apprentissage.

Les travaux de Lablidi et al. (2009), ainsi que ceux de Ezzahri et al. (2006), apportent également une contribution importante à l'évaluation des

plateformes de formation à distance. Ils ont analysé l'adéquation entre les potentialités technologiques de ces plateformes et les objectifs pédagogiques visés. Leur grille d'évaluation, intégrée à un environnement numérique d'aide à la conception et à l'évaluation des cours multimédias interactifs (CMI), prend en compte plusieurs dimensions : gestion de projet, qualité pédagogique, didactique et aspects techniques.

Ainsi, la recherche scientifique en évaluation au Maroc, bien qu'encore en voie de structuration, témoigne d'une dynamique riche, ancrée dans les réalités pédagogiques nationales et ouverte aux enjeux technologiques contemporains. Les travaux menés, portés par une diversité d'approches et d'acteurs, contribuent à redéfinir les contours de l'évaluation éducative, en la réorientant vers des finalités formatives, participatives et innovantes. Cette vitalité scientifique constitue un levier stratégique pour renforcer la qualité des dispositifs éducatifs et guider les réformes futures.

Dans les prochaines années, il pourrait être opportun que les chercheuses et chercheurs marocains portent leur attention sur :

1. La validation psychométrique de l'e-évaluation en contextes de massification (grands effectifs) : équivalence des épreuves, fiabilité des scores et invariance de mesure.
2. Les effets de l'IA et des learning analytics sur l'équité, la transparence, la protection des données et la traçabilité des décisions évaluatives.
3. Les modalités d'institutionnalisation de l'évaluation formative (gouvernance, référentiels, dispositifs d'accompagnement) et leurs conditions de pérennisation.
4. Les impacts à long terme des pratiques évaluatives sur la motivation, la persévérance et la réussite des apprenants.
5. La professionnalisation des enseignants à l'évaluation : compétences, instrumentation et éthique des données.
6. Les biais psychosociologiques et la qualité/ergonomie des plateformes : identification/réduction des biais (sévérité/complaisance, halo, affinité) et optimisation de l'expérience apprenant.

Ces orientations constituent des pistes de recherche prioritaires, à même d'éclairer les choix stratégiques et d'améliorer la qualité, l'équité et l'efficacité des dispositifs d'évaluation au Maroc.

V. Conclusion

L'évolution de la culture de l'évaluation en éducation et en formation révèle une dynamique à la fois complexe et prometteuse. Fondée sur un ancrage théorique structuré, elle s'est progressivement nourrie de réformes éducatives, d'initiatives institutionnelles et d'expérimentations pédagogiques, souvent portées par des acteurs engagés dans la transformation qualitative de l'enseignement. Différents intervenants ont, à travers divers dispositifs, tenté de promouvoir des pratiques évaluatives plus transparentes, plus formatives et plus équitables. Toutefois, ces ambitions se heurtent encore à des freins structurels et organisationnels.

Sur le terrain, l'évaluation demeure marquée par des tensions entre contrôle et régulation, normes prescrites et usages contextualisés, ou encore entre exigences institutionnelles et besoins des apprenants. Des expérimentations innovantes – telles que les tâches complexes, les feedbacks interactifs, les évaluations par les pairs ou l'intégration du numérique – témoignent néanmoins d'une volonté d'aligner les pratiques évaluatives sur les finalités d'un apprentissage durable, motivant et ouvert à la diversité des profils.

Parallèlement, les recherches académiques menées au sein de structures telles que l'ORDIPU et la section marocaine de l'ADMEE-Europe ont permis de documenter, d'analyser et de reconfigurer les logiques d'évaluation à partir de données empiriques. Ces travaux, de plus en plus diversifiés et ancrés dans les réalités locales, contribuent à repositionner l'évaluation comme un levier de transformation pédagogique, de pilotage stratégique et d'équité éducative.

Dès lors, il apparaît indispensable de consolider une culture évaluative inclusive fondée sur l'articulation entre recherche, formation initiale et continue, engagement des praticiens et pilotage institutionnel. Cette culture ne peut se construire sans un travail collectif, éthique et critique, capable de penser l'évaluation au-delà de sa fonction certificative, en tant que vecteur d'apprentissage, de réflexivité et d'émancipation.

L'enjeu, à l'ère du numérique et des compétences du XXIe siècle, est de transformer l'évaluation en un dispositif intelligent, évolutif et équitable, au service de la qualité de l'enseignement, de la réussite des apprenants et de l'amélioration continue des pratiques éducatives.

VI. Bibliographie

Akrim, H., Figari, G., Mottier-Lopez, L. et Talbi, M. (2010). La place de l'évaluation dans la réforme du système éducatif marocain : Questions pour la recherche. *Questions Vives, Recherches en éducation*, *4*(13), 387-400. https://doi.org/10.4000/questionsvives.323

Anajar, A., Talbi, M., Radid, M., Snadrou, K. et Tragha, A. (2015). Quality management in vocational training: Evaluation of a specialized institution in ICT. *Procedia - Social and Behavioral Sciences*, 191, 1928-1933. https://doi.org/10.1016/j.sbspro.2015.04.261

Ayat, H., Sadiq, M., Radid, M. et Chemsi, G. (2024). Revitalizing continuous education: Crafting a SPOC through social learning theories. Dans G. Chemsi, I. Elimadi, M. Sadiq et M. Radid (dir.), *Teaching and assessment in the era of Education 5.0* (p. 243-264). IGI Global. https://doi.org/10.4018/979-8-3693-3045-6.ch014

Bennani, M. (1986). Problématique de l'évaluation dans l'enseignement secondaire. *Revue marocaine des sciences de l'éducation*.

Chakir, A. (2022). Évaluation numérique dans l'enseignement supérieur au Maroc. *Revue internationale de pédagogie universitaire*, *38*(1), 55-72.

Chemsi, G., Sadiq, M., Radid, M. et Talbi, M. (2020). Study of the self-determined motivation among students in the context of online pedagogical activities. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(5), 17-29. https://doi.org/10.3991/ijet.v15i05.11392

Chemsi, G., Sadiq, M., Radid, M. et Talbi, M. (2016). Conception et analyse de la structure factorielle d'un outil informatisé d'évaluation des enseignements. *European Scientific Journal*, *12*(7), 109-123. https://doi.org/10.19044/esj.2016.v12n7p109

Dari, S., Aouad, S. A., Baha, F., Lajane, H., Gouifrane, R., Radid, M. et Chemsi, G. (2024). Evaluation and redesign of an e-learning platform. Dans G. Chemsi, I. Elimadi, M. Sadiq et M. Radid (dir.), *Teaching and assessment in the era of Education 5.0* (p. 115-134). IGI Global. https://doi.org/10.4018/979-8-3693-3045-6.ch007

El Gharbi, M. (2019). Pratiques d'évaluation critériée : Étude dans les lycées de la région Rabat-Salé-Kénitra [Thèse de doctorat, Université Mohammed V].

El Bousaadani, A. (2010). Biais psychosociologiques de l'évaluation par les acteurs de la mise en place d'une réforme [Thèse de doctorat, Université Hassan II de Casablanca].

El Machtani El Idrissi, W., Chemsi, G., El Kababi, K. et Radid, M. (2022). The impact of serious game on the nursing students' learning, behavioral engagement, and motivation. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(1), 18-35. https://doi.org/10.3991/ijet.v17i01.26857

Elimadi, I., Chafiq, N. et Talbi, M. (2024). Étude des effets des plateformes d'apprentissage en ligne Loghat et Moodle sur les expériences d'apprentissage des étudiants de la Faculté des sciences à l'ère numérique. Dans *Méthodologies et technologies d'apprentissage de l'enseignement supérieur en ligne* (p. 566-579). Springer. https://doi.org/10.1007/978-3-031-67351-1_37

Essaoudi, M., Lotfi, R., Lotfi, S., Radid, M. et Talbi, M. (2012, janvier). *L'évaluation des compétences professionnelles des étudiants inspecteurs de l'enseignement en fin de formation au Centre National de Formation des Inspecteurs de l'Enseignement (CNFIE) de Rabat (Maroc).* Communication présentée au 24e colloque international de l'ADMEE-Europe, Luxembourg.

Ezzahri, S., Talbi, M., Erradi, M., Khaldi, M. et Aride, J. (2006, novembre). *Élaboration d'un modèle de grille d'évaluation et d'analyses de produit pédagogique multimédia interactif en ligne*. Communication présentée au colloque CEMAFORAD 3 – Intégration des TIC à l'enseignement et à la formation : Réalités et perspectives, 13-15 novembre, Sousse, Tunisie.

Hajji, K. et Hamdani, S. (2017). Pratiques d'évaluation en contexte scolaire marocain : Entre discours et réalité. *Revue marocaine des sciences de l'éducation*, *12*(1), 45-62.

Hajji, S. (2016). Évaluation formative au secondaire : Perceptions et pratiques des enseignants. *Revue africaine de sciences de l'éducation.*

Hamdani, S. (2022). *L'évaluation formative dans l'enseignement : Enjeux et perspectives*. Centre Marocain de Recherche Pédagogique.

Housni, M., Namir, A., Talbi, M. et Chafiq, N. (2019). Applying data analytics and cumulative accuracy profile (CAP) approach in real-time maintenance of instructional design models. Dans

Lecture Notes in Real-Time Intelligent Systems (p. 17-25). Springer. https://doi.org/10.1007/978-3-319-91337-7_2

Khyati, A., Talbi, M. et Boumahmaza, M. (2013, janvier). *Former à l'évaluation des apprentissages en physique et chimie*. Communication présentée au 25e colloque de l'ADMEE-Europe, Fribourg (Suisse) : Évaluation et autoévaluation, quels espaces de formation. https://www.admee2013.ch/contributions/ADMEE-2013-Khyati-Talbi-Boumahmaza.pdf

Laaziz, Y., Chemsi, G. et Radid, M. (2024). The influence of e-assessment on students' cognitive engagement in higher education. *International Journal of Engineering Pedagogy (iJEP)*, *14*(4), 54-67. https://doi.org/10.3991/ijep.v14i4.48293

Lablidi, A., Abourrich, A. et Talbi, M. (2009, décembre). Démarche préconisée pour évaluer une plate-forme. *EpiNet – Revue électronique de l'EPI*, (120). https://www.epi.asso.fr/revue/articles/a0912b.htm

Lajane, H., Arai, M., Gouifrane, R., Qaisar, R., El Machtani El Idrissi, W., Chemsi, G. et Radid, M. (2021). A scenario of the formative e-assessment based on the ARCS model: What is the impact on student motivation in educational context? *International Journal of Emerging Technologies in Learning (iJET)*, *16*(24), 135-148. https://doi.org/10.3991/ijet.v16i24.24121

Lajane, H., Gouifrane, R., Arai, M., Qaisar, R., Chemsi, G. et Radid, M. (2023). Mobile technology for learning assessment: Design process for the NursE-Quiz mobile application. *International Journal of Interactive Mobile Technologies (iJIM)*, *17*(10), 228-241. https://online-journals.org/index.php/i-jim/article/view/37053

Maghnouj, S., Fordham, E., Clarke, J. et Kitchen, H. (2018). *Examens de l'OCDE du cadre d'évaluation de l'éducation : Maroc*. Éditions OCDE. https://doi.org/10.1787/9789264301832-fr

Moundy, K., Chafiq, N. et Talbi, M. (2021). Comparative analysis of student engagement in digital textbook use during quarantine. *Education Sciences*, *11*(7), 352. https://doi.org/10.3390/educsci11070352

Mrisse, F.-E., Moundy, K., Chafiq, N. et Talbi, M. (2023). Les pratiques d'évaluation en ligne dans un dispositif numérique en contexte de formation universitaire : Le cas de l'Université Hassan II.

International Journal of Advanced Computer Science and Applications (IJACSA), *14*(1). https://doi.org/10.14569/IJACSA.2023.0140148

Naciri, A., Radid, M., Sine, H., Kharbach, A. et Chemsi, G. (2023). Self-determined motivation and associated factors among health professions students in distance learning: A cross-sectional study in Morocco. *Korean Journal of Medical Education*, *35*(1), 33-43. https://doi.org/10.3946/kjme.2023.247

OCDE. (2024). *L'évaluation de la performance des établissements scolaires au Maroc*. Éditions OCDE. https://doi.org/10.1787/4f59bfc1-fr

Ouchaouka, L., Omari, K., Talbi, M., Moussetad, M., El Amrani, N. et Labriji, L. (2019). Flipped classroom and serious games as a new learning model in experimental sciences at the university. Dans M. E. Auer, H. Hortsch et P. Sethakul (dir.), *The Impact of the 4th Industrial Revolution on Engineering Education*. ICL 2019 (p. 731-739). Advances in Intelligent Systems and Computing, 1134. Springer, Cham. https://doi.org/10.1007/978-3-030-40274-7_69

Terhzaz, J., Akrim, H. et El Jamali, S. (2015, avril). *L'évaluation des enseignements par les professeurs stagiaires au CRMEF du Grand Casablanca*. Communication présentée à la Troisième rencontre nationale de didactique des sciences : L'évaluation pédagogique entre le discours scientifique et la mise en œuvre, Meknès, Maroc, 8-9 avril 2015.

Tilioui, C., Bellfkih, E. M., Imrane, C. I., El Kababi, K., Radid, M. et Chemsi, G. (2025). Exploring the impact of student orientation on mathematics learning using self-organizing maps: A study with middle school students. *Educational Process: International Journal*, *14*, e2025008. https://doi.org/10.22521/edupij.2025.14.8

Zoubaidi, W., Talbi, M., Lotfi, S., Akrim, H., Ihrachen, M., Leleu Galland, E., Figari, G. et Touri, B. (2014). Training of trainers of preschool educators: Analysis of needs and practices in the development of training device and skills referential. *Procedia – Social and Behavioral Sciences*, *116*, 1766–1773. https://doi.org/10.1016/j.sbspro.2014.01.470

Chapitre 20. Quarante ans d'évolution de l'évaluation au Liban : Enjeux, réformes et perspectives

Wassim El-Khatib, Patricia Azoury

I. Introduction

La question de l'évaluation des apprentissages est au cœur des politiques éducatives. Elle est en perpétuelle évolution, en fonction des exigences pédagogiques, sociales et politiques.

Au Liban, longtemps réduite à une fonction certificative à travers les examens nationaux – le Brevet et le Baccalauréat –, l'évaluation est devenue, en l'espace de quatre décennies, un enjeu stratégique.

Ce chapitre propose une lecture diachronique et critique de l'évolution de l'évaluation au Liban depuis les années 1980. Il s'intéresse à retracer le parcours à travers les réformes institutionnelles, les dynamiques pédagogiques et les influences internationales. Dans cette perspective, il mettra en lumière les défis persistants entre les normes prescrites et les pratiques effectives sur le terrain, tiraillées entre une volonté d'innovation et un attachement aux modèles évaluatifs traditionnels. En outre, ce chapitre analysera les politiques, les pratiques et les recherches dans le domaine de l'évaluation dans l'optique de comprendre les enjeux actuels de l'évaluation dans le contexte libanais.

L'analyse est structurée en quatre périodes :

- La première période : Les premières recherches en évaluation (1980-1990). Cette période est marquée par la domination des examens standardisés et l'absence d'une logique de l'évaluation au service de l'apprenant et de l'apprentissage.
- La deuxième période : Institutionnalisation et réforme (1990-2010). Cette période est caractérisée par une refonte curriculaire et l'émergence de nouvelles formes d'évaluation.
- La troisième période : Innovations et défis contemporains (2010-2020). Au cours de cette période, des pratiques innovantes d'évaluation émergent, entre autres, la coévaluation, l'autoévaluation et l'évaluation par le portfolio.
- La quatrième période : vers une politique renouvelée de l'évaluation (2020-2024). Cette période est marquée par une nouvelle refonte curriculaire et des perspectives prometteuses dans un contexte rongé par des crises multiformes.

Ce chapitre se veut donc une contribution à la compréhension des enjeux de l'évaluation dans le contexte libanais, tout en retraçant son évolution à partir d'un examen critique d'une littérature qui n'a cessé de se développer au fil de quatre décennies.

II. Méthodologie de l'analyse

La méthodologie adoptée est de type qualitatif documentaire. Elle a consisté à adopter une approche historico analytique s'inscrivant dans une logique compréhensive critique (Ricoeur, 1986 ; Ball, 2012). Cette approche cherche à attribuer une signification aux pratiques en analysant leur contexte social et en faisant apparaître leurs enjeux politiques, pédagogiques et culturels. L'évaluation, dans cette perspective, n'est pas considérée comme un acte neutre ou technique, mais s'inscrivant dans une logique pédagogique, institutionnelle et sociale (Perrenoud, 1998 ; De Ketele et Roegiers, 1993).

Le corpus est formé d'un ensemble de documents éducatifs, scientifiques et institutionnels portant sur l'évaluation au Liban entre 1980 et 2024. Il regroupe des textes de réforme, des rapports produits par les instances éducatives au Liban telles que le ministère de l'Éducation et de l'Enseignement Supérieur (dorénavant MEES) et le Centre de Recherche et

de Documentation Pédagogiques (dorénavant CRDP), des thèses, des mémoires de master, des articles parus dans des revues scientifiques spécialisées, des communications proposées lors de colloques.

Ces documents ont été sélectionnés dans des bases de données ouvertes telles que *Google Scholar, HAL, Thèses.fr, OpenEdition, Eric*, etc. et dans les bibliothèques des grandes universités francophones et anglophones suivantes :

- Université Saint-Joseph de Beyrouth (USJ),
- Université Libanaise (UL),
- Université Saint-Esprit de Kaslik (USEK),
- Université de Balamand,
- Université Américaine de Beyrouth (AUB).

En ce qui concerne les documents institutionnels, ils ont été consultés dans les archives du MEES et du CRDP. Ces documents incluent les instructions officielles, les *curricula*, les rapports de suivi et les publications de ces deux principales instances éducatives au Liban.

Pour cibler la recherche et orienter un choix pertinent des documents, une famille des termes clés a été établie. Les termes couvrent les dimensions diversifiées de l'évaluation telles que : évaluation formative, sommative, diagnostique, autoévaluation, coévaluation, critères d'évaluation, critère de réussite, critères de réalisation, mesure, évaluation authentique, et leurs équivalents en anglais.

Pour circonscrire, par ailleurs, la recherche au champ éducatif libanais, ces termes clés ont été croisés avec les termes suivants : Liban, contexte libanais, éducation au Liban et leurs équivalents en anglais bien entendu, tout en veillant à ce que la recherche couvre un éventail de types variés.

La grille d'analyse a été articulée autour des dimensions suivantes :

- les finalités de l'évaluation (certification, régulation, accompagnement),
- les objets évalués (savoirs, objectifs, compétences),
- les modalités et outils d'évaluation (examens, évaluation formative, coévaluation, outils numériques),
- les approches adoptées (pédagogie par objectifs, approche par compétences)

- et les défis rencontrés (résistances, contraintes systémiques, décalage entre prescriptions et pratiques).

III. Les premières recherches en évaluation (1980-1990)

Dans ce cadre temporel entre 1980 et 1990 - marqué par la fin de la guerre civile et l'essor des recherches en évaluation dans le champ des sciences de l'éducation, le Liban s'est inscrit, au niveau institutionnel, dans une logique d'évaluation spécifiquement certificative régie par les examens nationaux, nommément le Brevet et le Baccalauréat. Sur le plan de la recherche, quelques rares travaux commençaient à évoquer une nouvelle perspective de l'évaluation plus au service de l'apprenant et de l'apprentissage.

La terminologie de l'époque reflétait bien ces choix en matière d'évaluation. En effet, les termes utilisés pour désigner l'acte évaluatif se limitaient à « imtihān (examen) », « ikhtibār (épreuve) », « ālāmāt (notes) » et renvoyaient ainsi au modèle de « l'évaluation comme mesure » (Bonniol et Vial, 1997). L'évaluation, dans cette optique, s'intéresse à l'administration de la preuve, à l'objectivité, au « produit » beaucoup plus qu'au « processus » pour reprendre la distinction établie entre deux logiques ayant des visées différentes et relevant des paradigmes tout à fait distincts (Ardoino et Berger, 1989).

Le terme « taqwīm (تقويم) », équivalent arabe de l'évaluation au sens d'une démarche favorisant la régulation, était complètement absente de la littérature administrative de l'époque. Force est de constater ici qu'il ne s'agit pas d'un problème de vocabulaire. Le verdict prononcé dans le plan du redressement éducatif (CRDP, 1994) à l'égard des programmes en vigueur à l'époque, datant de 1968, est sans appel. Dans ce texte fondateur, le système éducatif est qualifié de « système dominé par l'obsession des examens officiels », qui « pêche par manque de prise en compte des critères d'évaluation modernes, à valeur scientifique ». Le jugement des « systèmes d'évaluation » n'est pas moins virulent : « ils évaluent les compétences de mémorisation et de reproduction des informations sans tenir compte de la capacité à mettre en application les connaissances acquises ni d'autres compétences de type intellectuel ou en rapport avec la créativité (résoudre des problèmes – analyser – créer – produire - inventer) ». Nous faisons

l'économie de l'évaluation des « auxiliaires pédagogiques » et des « méthodes d'enseignement », exprimée dans le même style.

Nonobstant le contexte institutionnel dominant, les premières recherches en matière d'évaluation qui voient le jour, remettent en question les limites et les biais de la logique de « contrôle » et prônent les fonctions formative, diagnostique et régulatrice de l'évaluation. Saigh (1990) s'intéresse aux différentes fonctions possibles de l'évaluation qui dépassent la fonction sommative et certificative et propose déjà une typologie des fonctions de l'évaluation.

Thoumy (1985), qui se penche, de son côté, sur l'évaluation dans le domaine de la formation des enseignants, appelle à une reconsidération des dispositifs adoptés et à l'articulation de ces derniers aux processus d'apprentissage et non aux principes de classement.

Ces initiatives, quoique sporadiques et limitées, ont ouvert la voie à un champ de recherche qui allait connaître un essor considérable. Elles ont préparé le terrain à la refonte des programmes qui allait se concrétiser dans le « Plan du redressement éducatif » ainsi que la parution des « Nouveaux programmes » (Décret-loi n° 10227 du 8 mai 1997), et ont jeté les bases, par ailleurs, d'une véritable inflexion de paradigme de l'enseignement vers celui de l'apprentissage.

IV. Institutionnalisation et réforme (1990-2010)

Après la fin de la guerre civile, qui a sévi pendant plus de trois décennies, il était devenu possible de passer à la restructuration du système éducatif dans la mesure où un tel chantier nécessitait forcément une entente sur le plan national. Les programmes parus, appelés « Nouveaux programmes », se sont alignés avec les grands principes pédagogiques de l'époque et ont affiché clairement une rupture avec les méthodes d'enseignement et d'évaluation dominantes jusqu'alors. Ces programmes s'inscrivaient délibérément dans les théories socioconstructivistes et rejetaient explicitement les conceptions behavioristes. Paradoxalement, lesdits programmes, malgré ces visées affichées, avaient toujours un fort ancrage dans l'approche par objectifs. La structure des programmes avec « finalités », « objectifs généraux » et « objectifs spécifiques » formait une illustration parfaite de la pédagogie par objectifs (PPO) telle que pensée par Mager (1962), Bloom (1979), De Landsheere, V. et De Landsheere G. (1988).

En matière d'évaluation, la confusion a été plus saillante. Le système d'évaluation prôné n'a pas figuré dans le décret-loi susmentionné, mais sous forme d'instructions postérieures introduisant une approche à laquelle les « Nouveaux programmes » ne font aucune mention, à savoir l'approche par compétences (APC). Ce manque d'alignement a été analysé par El-Khatib (2009) à la lumière des travaux de Vial (1997) sur le processus de référenciation, mettant en évidence des « difficultés à abandonner le modèle élu » et des « glissements non contrôlés » du « modèle de l'évaluation-régulation » au « modèle de la maîtrise par objectifs » (Vial, 1997, p.30).

Malgré ces incohérences, force est de constater que ces programmes ont eu la vertu d'introduire l'évaluation diagnostique et l'évaluation formative - inconnues jusqu'alors, d'amorcer une réflexion sur l'approche par compétences et de nourrir d'innombrables recherches sur tous les aspects liés à l'évaluation.

Certains travaux se sont intéressés au décalage entre les visées au niveau des programmes et les pratiques effectives sur le terrain. C'est ainsi qu'Abou Chacra (1998) a pointé des contradictions entre les choix curriculaires et leur application. Dans une veine similaire, Saliba (2008) a mis en lumière une prégnance des choix évaluatifs favorisant une évaluation formative au niveau des textes, mais une quasi-absence de cette forme d'évaluation au niveau de la pratique. El-Khatib (2002, 2009), de son côté, a mené une analyse sur les modèles de l'évaluation qui sous-tendent le système de l'évaluation préconisé en pointant les contradictions de ces choix. Il s'interroge également sur l'évaluation appropriée à la logique des compétences.

Par ailleurs, d'autres recherches, pendant cette période, ont porté sur le développement des dispositifs évaluatifs innovants. C'est le cas de Naccache et al, (2006) qui, dans le champ de la santé, ont montré à quel point le portfolio pouvait contribuer au développement de la réflexivité des étudiants dans le cadre de l'évaluation des compétences professionnelles. Dans la continuité de ces travaux, Bahous (2008) a conduit une recherche sur l'intérêt de l'usage du portfolio pour maintenir l'implication de l'apprenant et assurer son engagement dans le processus évaluatif. Les travaux de Halloun (2007) s'inscrivent dans cette lignée en mettant en évidence le rôle crucial des tâches complexes dans une approche en cohérence avec la logique de l'approche par compétences et la nécessité de former les enseignants et les responsables éducatifs pour assurer une implémentation efficace de cette approche.

Ce désir de promouvoir l'évaluation est entré en résonance avec des dynamiques portées par des instances internationales telles que l'UNESCO, la Banque Mondiale et l'Organisation internationale de la Francophonie (dorénavant OIF). C'est dans cette perspective que le programme IFADEM, lancé en 2007, a proposé des outils d'évaluation formative et a encouragé l'intégration du numérique dans les pratiques pédagogiques. Mais les rapports de suivi (CRDP, 2003) avec les données de l'étude PASEC (CONFEMEN, 2012) ont mis en évidence un décalage notable entre les désirs de changement et la réalité sur le terrain, en particulier dans les établissements publics. Cela est dû au manque d'accompagnement des enseignants du secteur public dans la mise en œuvre effective de l'approche par compétences. Au contraire, dans le privé, les dispositifs d'évaluation se sont plus développés grâce à l'autonomie dont ce secteur bénéficie : adoption de l'évaluation formative ou les bulletins fondés sur les compétences.

Sur un autre registre, petit à petit, une prise de conscience s'est manifestée quant à la nécessité de rendre l'évaluation plus inclusive, d'autant plus que certains établissements privés ont progressivement commencé à accueillir des élèves à besoins éducatifs particuliers (EBEP). La recherche de Salameh (2009) traduit parfaitement cette prise de conscience. En effet, l'auteur, dans sa quête d'un « référentiel national des compétences des enseignants de l'enseignement général au Liban », montre que les compétences relatives à la diversité des élèves et aux élèves à besoins éducatifs spéciaux des compétences s'avèrent essentielles à maîtriser.

Mais les progrès concrets sont restés limités en raison de l'absence d'un cadre législatif clair et du manque de ressources adaptées. Ainsi, cet intérêt croissant pour la question de l'évaluation a conduit plusieurs universitaires libanais à participer à des colloques spécialisés, tels que ceux organisés par l'ADMEE-Europe. Des communautés de recherche locales engagées dans une réflexion critique et comparative autour des enjeux de l'évaluation ont émergé sur le territoire libanais.

Enfin, bien que les bases et les modalités de l'évaluation aient été établies, la période 1990-2010 reste caractérisée par un paradoxe : d'un côté, les avancées notables réalisées sur le plan conceptuel sont accompagnées de diverses expérimentations ; de l'autre, des résistances d'ordre culturel, institutionnel et pédagogique ont freiné une transformation durable des pratiques.

En somme, ce parcours a retracé les pérégrinations d'une réforme prometteuse qui a ouvert la voie à de nouvelles réflexions sur l'évaluation des apprentissages.

V. Innovations et défis contemporains (2010-2020)

Décennie charnière dans l'évolution de l'évaluation des apprentissages au Liban, la période 2010-2020 s'est trouvée au confluent des injonctions à l'innovation et des inerties systémiques. Sous l'effet des appels internationaux à repenser les pratiques évaluatives et les dynamiques locales de formation, l'évaluation formative a vu sa place se renforcer dans les pratiques pédagogiques. Promue par les chercheurs et prônée par plusieurs projets institutionnels, elle s'impose désormais comme levier de régulation des apprentissages et d'accompagnement pédagogique.

Cette montée en puissance de l'évaluation formative, déclinée sous diverses formes (l'autoévaluation, la coévaluation, les grilles critériées et l'usage du portfolio) accentue sa dimension formatrice.

Les travaux sur les outils d'évaluation adaptés à la logique des compétences émergentes vont se multiplier. Dans cette perspective, les travaux de El Soufi (2017) analysent l'utilisation du e-portfolio dans un dispositif de formation à l'Université Libanaise, mettant en évidence son rôle dans l'autoévaluation, la construction des savoirs professionnels et l'engagement réflexif des futurs enseignants. Pour sa part, El-Khatib (2017) souligne l'importance des effets positifs du portfolio dans la formation des enseignants du secteur public, en tant qu'outil favorisant l'autorégulation et l'appropriation progressive des critères d'évaluation. Par ailleurs, Gérard (2010) défend la conception de situations d'évaluation authentiques engageant les apprenants dans des tâches complexes et contextualisées. Dans cette optique, plusieurs initiatives ont visé à repenser les modèles d'évaluation en élaborant des guides pédagogiques. L'un des exemples les plus marquants est le cours sur les méthodes d'enseignement et d'évaluation des matières d'études sociales (histoire, géographie, éducation civique), conçu par Bader et al., (2010) au sein du CRDP. Ce document opère un glissement significatif marquant le passage d'une logique psychométrique vers une approche intégrée de l'évaluation pédagogique.

Avec l'essor du numérique lors de cette période, un autre axe de la recherche a exploré l'impact du numérique sur le développement des pratiques évaluatives formatives et interactives. El-Soufi (2015), dont les travaux portent sur les usages de la plateforme Moodle, a observé que, malgré certaines résistances initiales, les dispositifs numériques permettent de mettre en place des formes d'évaluation formative et

interactive, en adéquation avec les attentes des étudiants. El-Saheli-El-Hage et Nahed (2020), dans l'objectif de tendre vers un contexte de transition numérique de l'éducation, ont jugé, de leur côté, important de fournir un feedback interactif et personnalisé afin d'accompagner l'apprentissage en ligne.

La recherche sur l'inclusion scolaire connaît ses prémices en explorant les pratiques évaluatives adaptées aux élèves à besoins spécifiques. Azoury et Zakaria (2023) ont développé des perspectives porteuses en considérant l'évaluation comme un outil puissant d'ajustement pédagogique dans le parcours d'inclusion. Elles ont toutefois mis en évidence les obstacles à une institutionnalisation de l'évaluation inclusive, comme la disparité des ressources et l'insuffisance de formation spécialisée. En outre, dans le contexte des écoles privées au Liban, les pratiques évaluatives à l'égard des élèves présentant des troubles d'apprentissage demeurent hétérogènes, oscillant entre des ajustements ponctuels et des dispositifs plus structurés (El-Saheli et al, 2016).

Face à la volonté constructive de faire évoluer les pratiques évaluatives à l'échelle nationale, des résistances persistent telles que la surcharge des programmes, la prédominance des examens standardisés et la pression sociale autour de la réussite certificative. Bien que de nombreux enseignants soient initiés aux nouvelles approches, leur application quotidienne s'avère difficile. Cette situation s'explique par le manque de temps, de ressources et l'absence d'une formation plus approfondie en évaluation formative (Riachi, 2015 ; Sarkis, 2016).

À ces difficultés structurelles sont venues s'ajouter les répercussions des crises socioéconomiques qui, depuis 2019, ont ébranlé profondément le système éducatif libanais. Les coupures d'électricité récurrentes, les grèves des enseignants dans le secteur public, les tensions politiques, la dévaluation de la monnaie, la détérioration des infrastructures scolaires : autant de facteurs qui entravent toute volonté de rénovation pédagogique. À cela s'ajoute la pandémie de la COVID-19 survenue en 2019 et qui a bouleversé le monde par la suite.

Ce contexte en pleine mutation a donné naissance au Liban, comme partout ailleurs, à des recherches articulées autour de l'enseignement en ligne et des défis que pose l'évaluation dans cette perspective. Les travaux de Rached et al., (2020) ont mis en lumière la bascule numérique imposée et ont souligné les lacunes du système vis-à-vis de la préparation techno-pédagogique.

Cette période a connu, en outre, un intérêt grandissant pour l'institutionnalisation de l'évaluation en créant un dispositif d'évaluation spécifique et autonome. Il s'agit du Dispositif National d'Évaluation (DNE) qui a vu le jour grâce à une collaboration entre le MEES et l'OIF (Organisation internationale de la Francophonie). Ce dispositif a permis de regrouper des experts issus des différentes instances éducatives au Liban : MEES, CRDP, Inspection pédagogique, DOPS (Direction d'orientation Pédagogique et Scolaire), mais aussi des experts internationaux. Bien que ce dispositif n'ait pas eu une longue existence, ce qui pourrait refléter en soi des défis et des tensions à l'égard de toute entreprise de rationaliser la politique de l'évaluation, il a permis d'offrir des indicateurs plausibles sur l'acquisition des compétences en français des élèves au terme de l'éducation de base. Grâce à une vaste enquête nationale, sous la coordination de Beacco (Beacco et al., 2016), le DNE a formulé des constats alarmants quant à l'acquisition des compétences évaluées et a mis en évidence les limites du système éducatif en matière d'acquisition des compétences de base, appelant à l'élaboration urgente d'une politique nationale d'évaluation formative et diagnostique tenant compte des réalités du terrain.

En parallèle, la structuration de la recherche scientifique en matière d'évaluation a franchi un cap important avec la création, en 2014, de la section libanaise de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE-Liban), à l'initiative du professeur El Hage. Regroupant des chercheurs provenant de plusieurs universités francophones, cette section s'est imposée progressivement comme un acteur clé dans la promotion d'une culture évaluative au Liban. À travers l'organisation de manifestations scientifiques, de séminaires de formation et de projets de recherche appliquée – souvent en partenariat avec l'Institut Français du Liban – l'ADMEE-Liban a contribué activement à la diffusion des pratiques innovantes en évaluation formative, au renforcement des compétences des praticiens et à l'élaboration d'une réflexion critique sur les politiques évaluatives dans le contexte libanais.

Pour conclure, en dépit des tensions soulevées, la décennie 2010-2020 demeure une étape charnière dans le domaine de l'évaluation pour avoir jeté les jalons d'une transformation profonde qui prendra corps dans les années suivantes avec le lancement d'une nouvelle refonte curriculaire et l'émergence de nouveaux axes de recherche en évaluation.

VI. Vers une politique renouvelée de l'évaluation (2020-2024)

Les années s'étendant de 2020 à 2024 constituent un tournant décisif dans l'évolution de l'évaluation au Liban. Les mutations profondes survenues à la suite des différentes crises économique, sanitaire, sociale et sécuritaire ont rendu le terrain propice à des changements structurels et au développement de nouvelles tendances de recherche en évaluation. Ce contexte est comparable à celui qui régnait à la fin de la guerre civile ayant permis une refonte curriculaire après trois décennies de stagnation. En effet, une fois encore, après presque trois décennies (1997-2022), le Liban procède à une nouvelle refonte curriculaire avec la publication du Cadre référentiel libanais du curriculum général pré-universitaire (CRDP, 2022).

Ce Cadre est la base de la réforme curriculaire entreprise et dont l'évaluation est une préoccupation majeure. Une réelle volonté de remédier aux incohérences du système d'évaluation des programmes en vigueur est affichée. Le Cadre met en avant la nécessité d'aligner l'évaluation sur l'approche par compétences visant le développement des compétences disciplinaires et transversales qualifiée « d'option éducative unificatrice et globale », et d'adopter des « méthodes et outils novateurs porteurs » tels que « l'autoévaluation », « l'évaluation par les pairs » et le « portfolio » qui permettent de garantir une « rétroaction personnalisée » pour l'apprenant.

Pour expliciter la politique et les démarches évaluatives à adopter, un document spécifique complémentaire a été réservé à l'évaluation. Ce document constitue l'un des onze « papiers » complémentaires du Cadre. Cette réforme est considérée par Nahas (2023) comme une rupture décisive avec les conceptions antérieures de l'évaluation principalement fondées sur la mémorisation et la sélection, même si son implémentation reste conditionnée par un changement radical de culture pédagogique. Abstraction faite de la validité de ce constat mettant en doute les entreprises antérieures visant à instaurer une évaluation au service de l'apprenant et de l'apprentissage, force est d'admettre que la nouvelle politique d'évaluation est perçue comme novatrice et relevant d'une logique tout à fait différente.

Ce changement fondamental de polarité dans le discours institutionnel sur l'évaluation suscite déjà une réflexion au niveau de la recherche, pour offrir des réponses scientifiquement validées sur la démarche à entreprendre pour évaluer les compétences transversales situées au cœur

de cette réforme et dont le développement et voire l'existence sont considérés parfois dans la littérature, comme problématique (Tardif et Dubois, 2013). Des communications aux colloques, des recherches en cours et des projets de collaboration entre équipes locales et à l'étranger sont envisagés. Ces différentes perspectives ont été rendues possibles grâce aux manifestations scientifiques et rencontres organisées par l'ADMEE-Europe.

Outre ce chantier engagé sur le plan national, des recherches sur l'enseignement et l'évaluation en ligne continuent à alimenter la réflexion sur cette problématique, d'autant plus que le basculement vers l'enseignement en ligne a révélé les failles structurelles du système éducatif libanais : les inégalités d'accès aux outils numériques, le manque de ressources pédagogiques disponibles et la faible préparation des enseignants à l'évaluation en ligne. Bitar (2023), Ghabash (2022), El Sibai (2022) et Mirza (2021) ont souligné, dans leurs recherches, que le passage à l'enseignement en ligne a accentué les écarts entre les élèves. Toutefois, les résultats de ces recherches s'accordent à considérer que ce basculement a permis une prise de conscience nouvelle de l'importance de l'évaluation formative, de la flexibilité des outils et de la personnalisation des retours sur les apprentissages.

Un autre champ de recherche a continué à se développer de manière accrue avec l'importance accordée à l'inclusion scolaire, qui forme, par ailleurs, une autre préoccupation majeure des nouveaux programmes. Les recherches d'Azoury (2022) se déploient dans cette perspective pour insister sur l'importance de modifier ces pratiques afin de répondre aux nouvelles exigences en matière d'évaluation des élèves à besoins éducatifs particuliers. Évaluer prend ainsi une nouvelle dimension, celle d'introduire des dispositifs d'aménagement approprié, de différenciation pédagogique et de reconnaissance explicite des divers profils d'apprentissage des élèves à besoins éducatifs particuliers et en situation de handicap. Ces approches innovantes restent peu généralisées dans le secteur public et privé. Cependant, leur inscription dans les textes curriculaires pose des jalons d'une évolution plus inclusive et systémique dans les années à venir.

En résumé, l'évaluation au cours de quatre années 2020-2024 a connu une accélération sans précédent en s'affirmant comme un des piliers essentiels de toute réforme éducative et en ouvrant des perspectives intéressantes à des champs de recherche innovants et porteurs.

VII. Conclusion

Cette analyse diachronique de 40 ans de recherche sur l'évaluation au Liban met en évidence que l'évaluation est bel et bien passée d'un champ de recherche incertain régi par une conception unilatérale pour devenir un champ fédérateur et porteur aligné sur les questions et les préoccupations actuelles. L'analyse montre que l'évaluation se trouve au carrefour des mutations éducative, sociale, sanitaire et économique et qu'elle constitue l'une des préoccupations majeures du système éducatif et de la recherche au Liban. Toutefois, force est d'admettre que la recherche sur l'évaluation au Liban n'a pas réussi à gagner en visibilité sur la scène de la recherche à l'international par manque de projets interuniversitaires locaux et internationaux et en l'absence des laboratoires spécialisés, sans négliger la carence en fonds de recherche étatiques ou institutionnels. Les recherches entreprises relèvent dans la quasi-majorité des cas d'initiatives personnelles isolées. Le contexte libanais, de par son infrastructure matérielle et académique avec une université publique et plus de 40 universités privées pour une population ne dépassant pas 6 millions d'habitants, est très propice au développement d'une recherche en évaluation de très grande qualité. Les pistes offertes sont, en effet, très porteuses. Dans un contexte national marqué par des crises permanentes où le recours aux dispositifs d'enseignement en ligne et ou hybrides s'impose, les recherches sur la validité et la fiabilité de ce type de dispositifs pourraient s'avérer pertinentes voire nécessaires. Les recherches pourraient également se tourner vers l'opérationnalisation des compétences transversales et leur évaluation pour nourrir une réflexion sur l'approche retenue dans la réforme curriculaire en cours, dont l'élément central est une approche par compétences visant le développement des compétences transversales et disciplinaires. Enfin, et sans prétendre épuiser les différentes possibilités, une piste importante concerne les évaluations internationales à l'aune des résultats jugés trop insatisfaisants du Liban aux tests de TIMSS et PISA.

VIII. Bibliographie

Abou Chacra, R. (1998). *Les nouveaux programmes d'enseignement du français au Liban : Intentions et tensions* [thèse de doctorat, Université de Franche-Comté]. https://theses.fr/1998BESA1016

Ardoino, J. et Berger, G. (1989). *D'une évaluation en miettes à une évaluation en actes*. Andsha-Matrice.

Azoury, P. (2022). *L'impact de l'approche par compétences sur le rendement scolaire des élèves à besoins éducatifs particuliers en langue française au niveau du langage écrit – Etude de cas dans les écoles publiques situées à Kesrwane et Jbeil - classe d'EB3*. [thèse de doctorat, Université Libanaise].

Azoury, P. et Zakaria, N. (2023, juillet). Évaluation de la formation dans le cadre de l'inclusion scolaire au Liban : De l'intervention éducative aux pratiques évaluatives. *Frantice.net, (21) Industries de la connaissance, éducation, formation et technologies pour le développement*, 21, Juillet 2023, (p. 33-48). http://frantice.net/index.php?id=1883

Bader, M., Khoury, S., Jefal, M. et Younès, J. (2010). *Cours sur les méthodes d'enseignement et d'évaluation des matières d'études sociales (histoire, géographie, éducation nationale et éducation civique)*. CRDP - Liban. https://tinyurl.com/32uvyh9h

Bahous, R. (2008). The self-assessed portfolio : A case study. *Assessing Writing*, *13*(4), 205–217. https://www.researchgate.net/publication/248966005_The_self-assessed_portfolio_A_case_stud

Ball, S. J. (2012). *Global Education Inc: New policy networks and the neoliberal imaginary*. Routledge.

Beacco, J.-C., Daou, H. et Chaib, L. (2016). *Évaluation nationale des compétences des élèves en français à la fin de l'éducation de base (EB9, ISCED2). Dispositif national d'évaluation du Liban (DNE).* CRDP et Institut français. https://tinyurl.com/4wn4kxdp

Bitar, S. E. (2023). *Évaluation des apprentissages à distance dans les conditions de confinement : Étude de cas dans une école privée au Liban-Nord* [thèse de doctorat, Université de Balamand]. https://scholarhub.balamand.edu.lb/handle/uob/6977

Bloom, B.S. et al. (1979). *Taxonomie des objectifs pédagogique*s. Tome 1 : Domaine cognitif (traduit par M. Lavallée). Les Presses de l'Université du Québec.

Bonniol, J.-J. et Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. De Bœck Université.

Centre de Recherche et de Développement Pédagogique (CRDP). (2022). الجامعي قبل ما العام التعليم لمنهاج اللبناني الوطني الإطار [Le cadre national libanais du curriculum de l'enseignement général préuniversitaire]. https://tinyurl.com/4uhe33s8

Centre de Recherche et de Développement Pédagogique. (1994). *Plan de relance éducative au Liban (Rapport n° EDU/94/1)*. Beyrouth. https://tinyurl.com/bdeep5hp

Centre de Recherche et de Développement Pédagogique. (2003). *Mise en œuvre de nouveaux programmes : problèmes, exigences et solutions proposées selon les opinions des directeurs des lycées publics*. Beyrouth. https://tinyurl.com/5xvbyvbw

CONFEMEN. (2012, 19 novembre). *Rapport du Liban provisoire*. https://tinyurl.com/2svzey

De Ketele, J.-M. et Roegiers, X. (1993). *Méthodologie du recueil d'informations*. De Boeck.

De Landsheere, V. et De Landsheere, G. (1988). *Définir les objectifs de l'éducation*. FeniXX.

El-Hage, F. et Nahed, R. (2020). Apprentissage par exploitation de l'erreur et à distance (AEED) : Pour une évaluation formative et un feedback interactif et digital. Évaluer. *Journal international de recherche en éducation et formation*, Numéro Hors-série, 1, 131–142. https://journal.admee.org/index.php/ejiref/article/view/179

El-Khatib, W. (2002). *L'évaluation dans les nouveaux programmes au Liban, entre les procédures de contrôle et l'évaluation au service de l'apprenant et de l'apprentissage*. [mémoire de DEA, Université Aix-Marseille 1].

El-Khatib, W. (2009). *Mise en place d'une approche par compétences en apprentissage de la langue française, dans le contexte libanais. Enjeux et difficultés*. [thèse de doctorat, Université Aix-Marseille I].

El-Khatib, W. (2017, décembre). Les effets pédagogiques de l'utilisation du portfolio comme outil d'évaluation dans la formation des enseignants du secteur public à la Faculté de Pédagogie de l'Université Libanaise. *Manarat Thaqafia,* (3), 176–202.

El-Saheli-El-Hage, R. et Sawilowsky, S. (2016). Assessment practices for students with learning disabilities in Lebanese private schools: A national survey. *Cogent Education*, *3*(1), 1261568.

El-Sibai, R. (2022). L'*évaluation des apprentissages à distance dans les écoles privées libanaises durant la pandémie de COVID-19* [mémoire de master, Université de Balamand].

El-Soufi, A. (2015). Utilisation de la plateforme Moodle : Analyse des perceptions des étudiantes de l'Institut libanais d'éducateurs. *Revue internationale des technologies en pédagogie universitaire*, *12*(3), 51–61. https://doi.org/10.18162/ritpu-2015-v12n3-04

El-Soufi, A. (2017). *Le e-portfolio dans la formation initiale des enseignants : un levier d'autoévaluation et de professionnalisation* [mémoire de DEA, Université Libanaise].

Gérard, F.-M. (2010). *Construire des évaluations au service des apprentissages*. De Boeck.

Ghabash, M. (2022). *L'évaluation des apprentissages à distance dans les écoles privées libanaises durant la pandémie de COVID-19* [mémoire de master, Université de Balamand].

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

Halloun, I. A. (2007). *Modeling theory in science education*. Springer.

Mager, R. (1962) *Preparing Instructional Objectives*. Fearon Publishers.

Mirza, H. S. (2021). University teachers' perceptions of online assessment during the Covid-19 pandemic in Lebanon. *American Academic & Scholarly Research Journal* (11-24), 11(10).

Naccache, N., Samson, L. et Jouquan, J. (2006). Le portfolio en éducation des sciences de la santé : Un outil d'apprentissage, de développement professionnel et d'évaluation. *Pédagogie Médicale*, *7*(2), 110–127. https://doi.org/10.1051/pmed:2006016

Nahas, G. N. (2023). Expérience d'une refonte curriculaire – Étude de cas : le Liban en 2022. *La Recherche en Éducation*, 28, 19–34.

Organisation de coopération et de développement économiques (OCDE). (2018). *The future of education and skills: Education 2030*. https://www.oecd.org/education/2030-project/

Perrenoud, P. (1998). *L'évaluation des élèves : De la fabrication de l'excellence à la régulation des apprentissages*. De Boeck.

Rached, P., Gharib, Y. et Constantin, S. (2020). Entre le cognitif et l'émotionnel... Quelle évaluation à distance en situation de crise ? *e-JIREF*, 223-230.

République libanaise. (1997). *Décret-loi n° 10227 du 8 mai 1997. Les curricula de l'enseignement général pré-universitaire et leurs objectifs*. https://www.crdp.org/curriculum1/173

Riachi, M. (2015*). Quelles conceptions et quelles pratiques des enseignants libanais en évaluation des apprentissages ?* [mémoire de master, Université de Balamand]. https://scholarhub.balamand.edu.lb/handle/uob/777

Ricoeur, P. (1986). *Du texte à l'action. Essais d'herméneutique*. Éditions du Seuil.

Saigh, P. A. (1990). School-based assessment research in Lebanon. McGill *Journal of Education/Revue des sciences de l'éducation de McGill*, *25*(1). https://mje.mcgill.ca/article/view/7899Typeset+2McGill Journal of Education+2McGill Journal of Education+2

Salameh, R. (2009, March). *Vers un Référentiel National des Compétences des Enseignants de L'enseignement Général au Liban*. Dans the Proceedings of The Professionalization of Teaching in Lebanon, Conference, French Version (p. 39-73). https://search.shamaa.org/PDF/43244Fr/SalameFr43259.pdf

Saliba, V. Z. (2008). *L'évaluation dans le nouveau programme de français au cycle moyen de l'éducation de base au Liban : Réalités et perspectives* [thèse de doctorat, Université Pierre Mendès France – Grenoble]. https://theses.fr/2008GRE29040

Sarkis, M. (2016). *Conceptions des enseignants en évaluation des apprentissages : Étude de cas dans une école privée à Jbeil (Liban)* [thèse de doctorat, Université de Balamand]. https://scholarhub.balamand.edu.lb/handle/uob/4439

Tardif, J. et Dubois, B. (2013). De la nature des compétences transversales jusqu'à leur évaluation : une course à obstacles, souvent infranchissables. *Revue française de linguistique appliquée*, *XVIII*(1), 29-45. https://doi.org/10.3917/rfla.181.0029

Thoumy, A. (1985). Esquisse d'un nouveau modèle pour la formation des maîtres de sciences au Liban. *Revue française de pédagogie*, 41-49.

Vial, M. (1997). Essai sur le processus de référenciation. L'évaluateur en habits de lumière. Dans J.-J. Bonniol et M. Vial, *Les modèles de l'évaluation, textes fondateurs avec commentaires* (p. 21-46). De Boeck et Larcier.

Chapitre 21. Du testing adaptatif à l'évaluation intelligente

40 ans de tissage entre le numérique et l'évaluation au sein de la Communauté ADMEE

Pierre-François Coen

I. Introduction

Le lien entre technologie et apprentissage n'est pas récent. Depuis les machines à enseigner chères à Skinner aux applications actuelles de l'IA en passant par l'apprentissage adaptatif ou la réalité virtuelle, l'espoir d'améliorer les performances des élèves, de comprendre leur manière d'apprendre ou encore d'individualiser l'enseignement a toujours été présent durant ces dernières décennies. Ainsi, que ce soit dans sa fonction formatrice ou certificative, l'évaluation est intimement liée à l'apprentissage. Il apparaît donc comme normal et évident que l'évaluation soit elle aussi très liée à la technologie. À la recherche de plus-value qualitative ou plus simplement pour gagner du temps, l'idée de technologiser l'évaluation est là depuis longtemps et a fait, elle aussi, un chemin impressionnant.

L'enjeu de ce chapitre est de voir comment la Communauté ADMEE a traité cette question. Nous allons donc retourner en arrière et faire un pas de côté pour prendre un peu de recul et voir comment le couple numérique - évaluation a évolué ces quarante dernières années. Pour guider cette

réflexion, nous allons nous appuyer essentiellement sur les travaux effectués au sein de la communauté de l'ADMEE en nous fondant principalement sur les écrits publiés dans les revues *Mesure et évaluation en éducation* (MEE) et *Journal international de recherche en éducation et formation* (E-JIREF). Dans cet exercice, nous ne revendiquons évidemment pas une analyse exhaustive de toutes les traces laissées durant toutes ces années (nous aurions pu aussi examiner tous les programmes des colloques de l'ADMEE), mais après un examen minutieux d'une cinquantaine d'écrits, il nous semble pertinent de donner quelques orientations susceptibles d'enrichir les réflexions actuelles. Précisons encore que nous nous sommes centrés sur des écrits significatifs à nos yeux qui prenaient pour objets le numérique. Nous avons délibérément laissé de côté ceux qui touchaient au numérique de manière indirecte, par exemple par le biais de méthodologies intégrant des outils numériques comme la vidéo.

Nous commencerons par donner quelques points de repères historiques et terminologiques. Puis, le chapitre se découpera en cinq sections qui croiseront à la fois des thématiques emblématiques et des balises temporelles. Nous terminerons par quelques constats conclusifs.

II. Quelques points de repère

Disons d'emblée que le couple évaluation - informatique est relativement peu présent dans la communauté ADMEE. En 1991, les rédacteur·trice·s de la revue MEE présentent une bibliographie thématique des 14 premiers volumes de la revue (de 1978 à 1991) et dénombrent relativement peu d'articles consacrés à ce thème. Comparativement aux autres thèmes, cette catégorie reste très marginale. En observant les publications de la revue MEE sur une plus longue période (voir figure n° 1), on peut dire qu'en moyenne, un article traitant de l'informatique/l'ordinateur/le numérique et l'évaluation est publié à peu près tous les deux trois ans. La cadence est à peine supérieure au début des années 90 qui coïncident avec la démocratisation des P.C. En 1990, un premier symposium autour du rôle des ordinateurs et de l'évaluation est organisé et aborde principalement des questions autour du testing adaptatif et diagnostique. Au fil des ans, les articles traitant des thématiques de validation des tests, de banque d'items, puis de numérique et dispositifs d'apprentissage (plateforme e-learning) s'égrènent encore timidement. En 2019, un symposium au colloque de 2018, conduit à la publication d'un numéro d'E-JIREF consacrée au numérique (Coen et al. 2019) et, une année plus tard, un opus spécial consacré à l'évaluation durant la pandémie du COVID permet le

partage de nombreuses pratiques d'évaluation en particulier dans des contextes d'enseignement à distance.

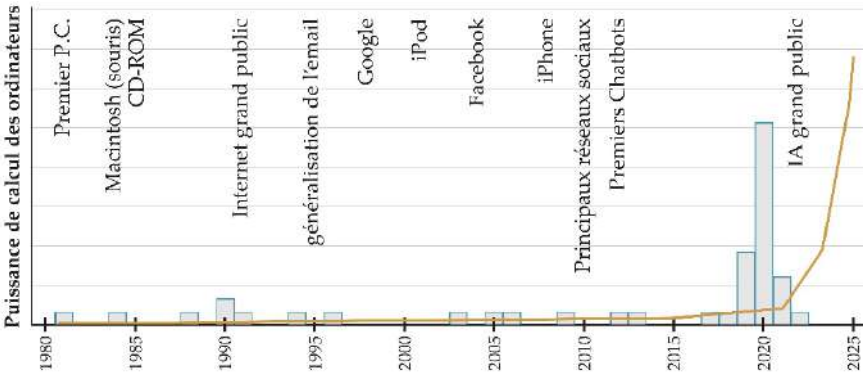


Figure 1 : Nombre de publications consacrées au numérique et à l'évaluation avec la courbe explicitant la puissance des ordinateurs ces 50 dernières années

Lorsqu'on parcourt les textes, on peut voir que différents termes ont été utilisés pour désigner ce qu'on appelle aujourd'hui les technologies numériques : informatique, ordinateur (micro-ordinateur) et évaluation, système informatisé, gestion informatisée de l'enseignement et des apprentissages, évaluation assistée par ordinateur, logiciels d'évaluation, NTIC (nouvelles technologies de l'information et de la communication) ou TIC (technologies de l'information et de la communication) et évaluation, apprentissage adaptatif, évaluation et intelligence artificielle, etc. Cette diversité montre une grande mouvance du champ et une certaine instabilité. Elle n'est pas surprenante et résulte sans doute de l'évolution très rapide des technologies dans la société durant ces cinquante dernières années. La figure n° 1 montre également le développement important (en particulier ces 5 dernières années) de la puissance de calcul des ordinateurs. Selon certains analystes, cette puissance a doublé quasiment tous les deux ans depuis 1960 et depuis l'avènement de l'IA, les choses se sont encore grandement accélérées (Connes, 2022) et touchent de plein fouet les pratiques d'évaluation ; nous y reviendrons.

Examinons maintenant quelques thématiques saillantes issues de la communauté ADMEE. Nous avons pris le parti de faire a priori une présentation chronologique des thématiques, même si cette règle ne sera pas toujours strictement suivie.

III. Les premiers pas du testing adaptatif

Le premier article explicitant clairement le rôle de l'ordinateur dans les pratiques d'évaluation est celui de Seguin (1984) intitulé *L'utilisation des micro-ordinateurs pour l'évaluation des apprentissages : quelques perspectives d'avenir*. L'auteur commence par dire que si l'ordinateur est utilisé comme support pour les apprentissages, il doit l'être également pour les évaluer. Il poursuit sa réflexion en évoquant une nouveauté importante pour l'époque : le testing adaptatif. Il note ainsi tout l'intérêt d'utiliser l'ordinateur pour constituer des banques d'items dont la passation serait grandement facilitée via l'ordinateur. Le paramétrage des items en termes de difficulté ainsi que « la conservation et l'analyse des résultats obtenus pour vérifier la validité de chaque item par une technique d'analyse scientifiquement éprouvée et reconnue » (p. 55) serait un atout indéniable de même que la possibilité de réduire la charge administrative (impression des notes, correction d'examens, mise à jour des dossiers cumulatifs). Pour lui, l'usage de l'ordinateur pourrait être fait dans une « évaluation formative par étapes itératives » (p. 60) au cours de laquelle l'élève répond à des questions et les réponses qu'il donne permettent de situer son niveau. La complexité de la démarche serait grandement facilitée par l'utilisation de l'ordinateur et conduirait petit à petit à ce que l'auteur appelle le testing adaptatif. Il termine par un plaidoyer intéressant sur « la gestion informatisée de l'évaluation pédagogique soutenant la mise à disposition de logiciel pour les enseignants pour leur faciliter l'accomplissement des tâches relatives à l'évaluation autrement longues et fastidieuses » (p. 60). Cette question du testing adaptatif va occuper les chercheurs durant de nombreuses années et jusqu'à aujourd'hui encore. En effet, l'idée d'ajuster le niveau des questions aux besoins des élèves est un élément clé de l'apprentissage soutenu par le numérique. On notera encore que ce thème est souvent associé à la notion de test diagnostique informatisé. C'est ainsi que les travaux de Dassa (1988), Dassa et De Cotret (1994), Laurier (1996), Dechef et Laveault (1999) aborderont cette question sous différents angles. En 2003, Martin (2003) fait une bonne synthèse des recherches dans le domaine, mais, dans son bilan, il fait état d'une certaine déception liée en partie à la difficulté de gérer la complexité et les nombreux paramètres de l'apprentissage adaptatif. Grâce à la puissance des outils actuels, le testing adaptatif est de plus en plus à même de répondre aux besoins spécifiques des élèves. Un engouement important est aujourd'hui à nouveau visible autour de ce thème.

IV. L'évaluation assistée par ordinateur (EAO)

Il faut attendre le passage des années 2000 pour voir apparaître le terme d'EAO (évaluation assistée par ordinateur) dans un texte de Dionne (2006). Cet écrit est intéressant parce qu'en dépit d'un parti pris assumé pour l'EAO, il apporte pour la première fois des éléments solides et très documentés sur le mariage entre ordinateur et évaluation. L'auteur commence par expliciter clairement les liens entre les tests et l'informatique et ses plus-values. D'abord, l'EAO permet une meilleure différenciation et un meilleur respect des rythmes d'apprentissages des élèves. Ensuite, il est possible d'ajuster l'épreuve en fonction du niveau de compétences de l'élève. Dionne ajoute qu'un « des avantages manifestes de l'EAO réside dans le temps économisé lors de la correction des épreuves. (...) Un environnement informatique permet de réaliser la correction et la compilation des résultats en un temps minimal » (p. 104). Enfin, il ajoute que « dans un proche avenir, il sera possible de traiter des items à réponses élaborées grâce aux développements issus de la recherche en correction automatique de rédactions écrites » (p. 104). Même si cette possibilité n'apparaîtra que plus tard, ce texte vante aussi l'EAO pour ses capacités à fournir des feedbacks immédiats, ce qui permet aux élèves et aux enseignant·e·s d'être renseignés sur la progression de l'apprentissage. Cela permet également de réduire le risque d'erreurs de correction et par une analyse des résultats, de cerner plus rapidement et facilement les difficultés des élèves de manière à proposer des remédiations plus adéquates. Des risques de tricheries sont évoqués par l'auteur qui note encore que l'EAO permet d'enrichir les items proposés aux élèves (recours à des situations avec des simulations pour « plonger les élèves dans une situation d'évaluation authentique » (p. 108) ou d'utiliser des laboratoires virtuels ou encore des extraits vidéo auxquels des questions sont associées). Par contre, l'auteur - fidèle aux principes fondateurs de l'ADMEE - souligne la nécessité de s'assurer de la validité des items. Cette tâche, pas si simple à réaliser, nécessite de s'astreindre tout autant que la construction de banque d'items rigoureusement élaborés en s'appuyant sur les réponses des élèves et, quand cela est possible, sur des données liées au processus de résolution de la question. L'article aborde pour terminer quelques aspects très intéressants comme les différences de validité entre les épreuves papier-crayon et électroniques, les facteurs d'appréhension liés à l'utilisation de l'ordinateur (en particulier entre filles et garçons) et des questions liées aux coûts d'implantation de ces outils (en précisant également les économies que cela pourrait amener en termes d'impression de papier).

Cette publication a été suivie de quelques autres abordant différents thèmes (Schuessler, 2009 ; Dionne, 2012 ; Béland, 2013 ; Morin, 2017 ; Drijvers, 2018 ; Sacré, 2019) qui s'égrènent jusqu'à la fin de la deuxième décennie des années 2000 où un numéro spécial de E-JIREF (*Evaluation et numérique : des pratiques éclectiques qui explorent des espaces à déchiffrer*) est consacré à cette thématique (Coen et al. 2019). Ce recueil, consécutif à un symposium organisé dans le cadre du colloque 2018 de l'ADMEE, touche de nombreux aspects comme la nature des données collectées et les modalités de récolte, les apports du numérique sur la régulation des apprentissages (diagnostic, gestion de la progression …), la nature des feedbacks et leurs impacts sur les apprenants, la diversité des objets évalués ainsi que des aspects éthiques. Dans leur introduction, les auteurs soulignent la grande hétérogénéité des articles qui démontrent des choix très différents dans les angles d'attaque des contributeur·trice·s. La grande ouverture thématique de l'ouvrage se vérifie dans les papiers qui touchent autant l'engagement cognitif des étudiants lorsque l'évaluation se passe avec de très grands groupes d'étudiants (Lepage et al. 2019), les apports technologiques dans un dispositif d'évaluation alliant des acteurs sur le terrain munis d'oreillettes et des spectateurs distants interagissant avec eux (Gardies et al. 2019), la comparaison de la quantité et la qualité des régulations entre classe réelle et classe virtuelle (Carrupt et Barras, 2019) que l'impact de l'approche par compétence, la formation à distance et l'utilisation des technologies (Leroux et al. 2019) ou encore l'approche par projet (David et Droyer, 2019) dans les dispositifs d'évaluation. À ce stade, on peut rejoindre le point de vue des auteur·e·s qui soulignent plusieurs points : d'abord le foisonnement de travaux dans ce champ et peut-être un certain manque d'axes structurants, l'amenuisement de la vague numérique observée dans le monde scolaire dès les années 2005, ensuite la résurgence de certains modèles d'évaluation « traditionnelle » renforcée par l'usage de plateformes présentant des outils relativement peu ouverts, et enfin, une ouverture vers l'utilisation plus systématique des *big data* pour permettre un retour aux « questions historiques » comme le diagnostic automatique, le profilage des apprenants, le testing adaptatif ou encore le pilotage des systèmes éducatifs par les chiffres, l'allocation des ressources fondées sur les algorithmes…

V. Evaluation formative, processus et hybridation de l'évaluation

L'évaluation formative et la régulation des apprentissages ont mobilisé un nombre très important de chercheur·euses de la communauté ADMEE. Il n'est donc pas surprenant de voir cette thématique associée à celle des technologies. En 1990 déjà, Boulet et al. conduisent une recherche intitulée *Système informatisé pour l'évaluation formative* dans laquelle l'équipe de recherche croise des contenus musicaux avec une architecture informatique permettant à l'enseignant·e de paramétrer finement les apprentissages de ses élèves. Quelques constats sont faits, en particulier sur les avantages d'obtenir des feedbacks immédiatement et de diminuer considérablement le temps de correction en dépit d'un temps plus grand pour répondre aux questions et de problèmes d'accès aux salles informatiques.

Dans ses travaux, Campos (1990), considère « qu'un des atouts majeurs de l'enseignement assisté par ordinateur est qu'il ouvre grand les portes d'une individualisation de l'apprentissage » (p. 56). Il doit permettre d'aller plus loin qu'un simple feedback de type réponse correcte/incorrecte en proposant des remédiations spécifiques aux types d'erreurs. L'auteur évoque dans son papier différentes recherches qui s'attachent à analyser les procédures erronées des élèves, mais laisse le lecteur un peu sur sa faim quand, après de grandes démonstrations sur le nombre de procédures à prendre en compte dans l'apprentissage de la soustraction, il conclut qu'« il est probablement vain de vouloir créer un didacticiel qui tenterait d'apporter une remédiation propre à chaque type d'erreurs » parce que « de l'enseignant et l'ordinateur, le premier est le plus flexible et le plus apte à moduler les remédiations en fonction des difficultés de l'élève. Le second est moins vite écœuré par le dépistage des erreurs et peut générer des exercices qui permettront de déceler la persistance de procédures erronées » (p. 65). Pour lui, il faut que la remédiation ait du sens et que « l'utilisation de tels outils (...) ne saurait être que profitable, dans la mesure où ils les [les enseignant·e·s] aident à se rendre compte que bien des erreurs ne sont pas dues à une simple inattention de l'élève, mais bien à une connaissance imparfaite de la matière abordée ». (p. 65). Constat qui sur le fond est déjà bien intéressant.

En matière d'évaluation formative, nous retiendrons également l'expérience originale de deux enseignants suisses (Favre et Briguet) qui, à la suite d'une conférence de Samuel Roller, développent dès 1991 un

dispositif informatique pour gérer la progression de leurs élèves. Ces deux praticiens créent le logiciel Evalog (https://evalog.ch/) qui permet de noter très précisément les objectifs maîtrisés (ou non) par les apprenants. Les fonctions du logiciel permettent de suivre ainsi les apprentissages et, en effectuant des tris et des regroupements d'élèves selon leurs besoins, de mettre en place une différenciation pédagogique efficace. Notons qu'après de nombreuses mises à jour, le logiciel est toujours opérationnel et adapté au plan d'études suisse.

Dix ans après ces travaux, Coen (1997, 2020) publie une thèse sur l'utilisation du logiciel AutoéVal qui permet à l'enseignant et à l'élève d'accéder au processus d'écriture. L'idée de base rejoint celle précédemment évoquée d'une prise de conscience de sa propre activité pour la réguler le mieux possible. Présentée au colloque de l'ADMEE de Mons en 1998, cette recherche explore quelques pistes intéressantes. D'abord celle d'utiliser l'ordinateur pour révéler ce qui n'est a priori pas accessible pour l'apprenant. Le logiciel produit un graphique mêlant la courbe de progression du texte à l'utilisation des outils à disposition qui sert ainsi de support pour permettre aux élèves d'expliciter leur stratégie d'écriture. Ensuite, l'analyse du protocole permet d'affiner considérablement le diagnostic des difficultés puisque l'enseignant·e peut accéder aux différentes régulations effectuées par le scripteur. Et enfin, par l'ajout d'outils spécifiques et personnalisés, il permet de donner à l'élève des aides à l'écriture parfaitement adaptées à ses besoins.

Avec le développement d'Internet, le recours à des logiciels particuliers s'amenuise et l'ordinateur fait place aux plateformes numériques qui intègrent souvent des dispositifs d'évaluation ou de régulation des apprentissages. Dans cette logique, on retiendra l'article de Dionne (2012) consacrés aux systèmes à réponses personnalisées (SRP). Ces dispositifs permettent de sonder les étudiants durant les cours via un boitier, un smartphone ou un ordinateur. Ils se présentent très souvent sous forme de quiz auxquels les étudiants doivent répondre pour donner aux enseignants les moyens d'évaluer leurs pré-acquis ou plus généralement la compréhension d'une séquence d'apprentissage. Dionne démontre que les retours donnés par les élèves sont des informations très intéressantes en particulier parce qu'elles sont immédiates et complémentaires à ce qu'il pourrait percevoir. Elles permettent de rétroagir et de proposer des ajustements de l'enseignement susceptibles de favoriser les apprentissages. Dans une revue de la littérature sur le sujet et ciblée sur l'enseignement supérieur, Detroz et Younès (2023) présentent différentes recherches récentes qui montrent l'intérêt de telles démarches pour favoriser les apprentissages profonds même si ces auteurs font état de

résultats contrastés mettant en évidence l'influence de différentes variables notamment le niveau de développement pédagogique des enseignant·e·s concerné·e·s.

Cette idée est prolongée par Alvarez et al. (2021) qui proposent une plateforme numérique (www.oura2.ch) permettant de créer des micro-sondages au fil des cours ciblant - non pas les apprentissages des étudiant·e·s eux-mêmes - mais des variables pédagogiques liées aux dispositifs proposés telles que l'attrait des tâches, leur valeur perçue, l'engagement des apprenants, leur sentiment de compétence, etc. L'idée centrale de cette démarche (utilisée notamment en formation) est que, dans une logique de contributions réciproques (Coen et al., 2024), les avis des élèves permettent aux professeurs d'améliorer leurs pratiques d'enseignement et de se développer professionnellement pour mieux répondre aux besoins des apprenant·e·s.

On notera enfin l'article de synthèse de Sacré et al. (2019) sur l'impact de composantes pédagogiques des plateformes numériques sur les apprentissages ou, dans un domaine plus spécifique, le travail de Loignon (2021) sur le développement d'un outil d'évaluation de la complexité linguistique.

VI. Le boum de la pandémie

Pour terminer ce petit panorama, il est évidemment impossible de passer sous silence l'impact très important de la pandémie du COVID-19. En matière d'éducation, le virage a été brutal avec l'obligation rapide d'utiliser massivement les technologies pour l'enseignement à distance lors du confinement. La communauté ADMEE n'a évidemment pas échappé à l'événement et un numéro spécial d'e-JREF a été consacré à cette période. Sans entrer dans une revue systématique des contenus de toutes ces publications, plusieurs thèmes se dégagent de ce numéro spécial et touchent de près ou de loin l'évaluation. Le premier - très représenté - aborde les nouveaux défis liés à l'évaluation à distance et les adaptations ou les transformations des dispositifs dans de nouveau contexte mobilisant de l'évaluation formative ou sommative (Burguete et al., 2020 ; Da Costa Cabral et al., 2020 ; Duroisin, 2020 ; El Hage et Nahed, 2020 ; Gilles et Charlier, 2020). Le deuxième met en évidence l'intérêt des technologies pour réduire la distance et ainsi accompagner de près la progression des apprenants en essayant de capter des traces de l'avancement de leur travail et en déployant des moyens pour les rassurer

(communication asynchrone) (Coen et Pellegrini, 2020 ; El Hage et Nahed, 2020 ; Thual, 2020). Plusieurs auteurs évoquent des aspects éthiques de l'évaluation médiatisée avec les technologies en termes de stress pour les élèves ou de relations avec les enseignants (Alonso Vilches et al., 2020 ; Da Costa Cabral et al., 2020 ; Lison, 2020). Enfin, par le biais de critères de qualité ou de questionnements ouverts, quelques textes se regroupent autour de recommandations concernant l'usage du numérique pour évaluer les apprentissages (Gilles et Charlier, 2020 ; Heilporn et Denis, 2020 ; Papi et al., 2020 ; Rached et al., 2020).

Avant d'aborder le dernier point, il nous semble nécessaire d'évoquer à ce stade l'ouvrage de Loye et Duroisin (2024), certes publié en marge de la revue *Mesure et évaluation*, mais qui regroupe des contributions de chercheur·euse·s très actif·ve·s au sein du réseau de l'ADMEE (à commencer par les auteures du livre). Abordant successivement les questions liées aux contextes de l'évaluation, à la collecte des données, à leur analyse et leur modélisation, les coordinatrices de l'ouvrage donnent à voir un renouvellement du champ par le foisonnement de thématiques abordées.

VII. Les enjeux de l'intelligence artificielle

Depuis l'arrivée dans le grand public de ChatGPT et des autres modèles de langage, de nombreuses questions se posent en termes d'apprentissage et d'évaluation. À tel point que l'IA semble écraser toutes les autres problématiques. Le dernier colloque de 2025 (Luxembourg) a vu exploser le nombre de contributions traitant de ce thème. Alors aujourd'hui, peut-on dire que les avantages et les risques liés à l'utilisation de l'IA dans l'évaluation des apprentissages s'équilibrent ? On peut certes soutenir que l'IA permet de gagner un temps précieux, de donner des feedbacks plus rapidement, plus personnalisés et plus adaptés, d'analyser plus finement les données liées au processus d'apprentissage et ainsi de mieux comprendre les erreurs et les difficultés des élèves. Mais on peut aussi y voir des risques importants de perte de sens des apprentissages pour les élèves, de manque d'adaptation face à des tâches moins standardisées, de manque d'empathie et de déshumanisation des interactions humaines ou encore de captation malveillante de données personnelles. Le dernier article dont nous parlerons ici aborde cinq enjeux d'évaluation face à l'émergence des IA génératives en éducation (Audran, 2024). En partant du principe que les « IA génératives obéissent à une logique qui a l'ambition de substituer une production numérique artificielle à la création

humaine », Audran pose, pour commencer, la question des compétences des utilisateurs et de leur capacité à comprendre le fonctionnement de ces outils. À la vue du nombre de paramètres mobilisés dans ces systèmes, il semble que « la seule opération évaluative que peut réaliser l'utilisateur final est d'estimer la qualité de la production à l'aune de ses propres connaissances » (p. 14). Le deuxième enjeu consiste à évaluer les illusions fondées sur la forme. En d'autres termes, il s'agit ici d'apprendre aux utilisateurs à « déceler les indices qui caractérisent les productions de ces générateurs et faire la différence entre production humaine originale et production machinique » (p. 15). Le troisième défi consiste à évaluer la part des hallucinations pour s'assurer que ce qui est produit est vrai. Cette question est cruciale et renvoie à des enjeux philosophiques (qu'est-ce que le vrai ?). Cela dit, cette aptitude reste indispensable à développer chez les utilisateurs en leur donnant les moyens de cette vérification. La dimension éthique est elle aussi très importante parce que le recours à ces outils permet à certains élèves de réaliser des tâches parfaitement conformes à ce qui est attendu. Il s'agit alors pour les enseignants d'évaluer l'authenticité des productions des élèves pour déceler d'éventuelles tricheries. Cette problématique n'est pas nouvelle, mais là, elle se pose de manière plus importante parce que sa détection est très difficile. Cette question en ouvre d'autres sur la complémentarité entre l'élève et l'IA dans la réalisation d'une tâche. Le dernier enjeu touche l'application (ou l'intégration) de l'IA pour améliorer l'évaluation. Comme nous avons pu le voir dans ce texte, l'idée de disposer d'outils capables de s'ajuster au niveau des élèves, de leur donner des tâches adaptées à leurs besoins habite les chercheurs depuis de nombreuses années. L'IA offre des perspectives intéressantes en permettant un affinage du diagnostic ou en assistant l'élève dans son autoévaluation. Dans ce sens, et à la suite d'Audran, il nous semble important de penser une intégration réfléchie, très réfléchie, de ces outils plutôt que d'organiser une résistance servile.

VIII. Constats conclusifs

Au terme de ce petit périple autour de l'évaluation et du numérique dans la communauté ADMEE, il nous semble intéressant de faire quelques constats conclusifs.

L'intérêt modéré des chercheurs pour lier évaluation et technologie ne traduit pas une volonté de mise à l'écart. L'examen des articles parus depuis plus de quarante ans nous montre au contraire une volonté d'associer avec prudence le numérique et l'évaluation. Le souci de garantir

la validité des tests et de comprendre comment fonctionnent les outils semble être encore très présent dans la communauté des chercheur·euse·s de l'ADMEE. Certes, on peut voir quelques décalages entre les avancées techniques et leur utilisation pour évaluer, mais cela nous semble assez normal en regard du lien qu'entretiennent depuis toujours l'école et la société.

Nous faisons aussi le constat de l'accroissement d'une certaine diversité dans les thématiques touchant l'évaluation et la technologie. Le numéro d'E-JIREF de 2019 le montre bien. Y aurait-il un éparpillement, une dispersion excessive des intérêts, des thématiques ? Difficile de le dire. Si le testing adaptatif occupait clairement la scène dans les années 90, la diversification des outils et des usages du numérique a conduit les chercheurs à explorer des champs nouveaux, sans qu'il y ait a priori des axes forts. C'est peut-être le propre de ce domaine d'études qui par la multiplicité des usages possibles des technologies rend difficile la cohérence et l'intelligibilité.

Enfin, l'arrivée de l'IA donne l'impression d'un écrasement, comme si cette technologie s'immisçait partout jusque dans les moindres interstices : diagnostic des difficultés des élèves, remédiations et production d'exercices adaptés, jugement évaluatif, production de critères, autoévaluation et suggestion d'améliorations, analyse de données et de l'activité, profilage selon des compétences, aide à la décision… Comme le rappelle Sadin (2015), après l'âge de l'accès et celui de la mesure, nous sommes désormais dans celui de l'injonction dans lequel les machines prescrivent les comportements à suivre. Serons-nous dès lors en mesure de garder notre marge de manœuvre, de rester maîtres du jeu et de cultiver notre esprit critique. Qu'adviendra-t-il des enfants qui n'apprendront pas comme nous à écrire, à analyser ou à créer « à l'ancienne » ? Ce regard sur le passé devrait à la fois nous rassurer, mais aussi nous ouvrir les yeux sur tous les possibles… autant les pires que les meilleurs pour toujours garantir aux élèves une évaluation intelligente et à visage humain.

IX. Bibliographie

Alonso Vilches, V., Detroz, P., Hausman, M. et Verpoorten, D. (2020). Réception de la prescription à « basculer vers l'eLearning » en période d'urgence sanitaire – Une étude de cas. *Journal international de recherche en éducation et en formation*, hors-série : évaluer en temps de pandémie, 5-16.

Alvarez, L., Çuko, K., Boéchat-Heer, S. et Coen, P.-F. (2021). Faciliter l'autorégulation de l'enseignement. Conception d'un dispositif numérique fondé sur les analytiques de l'enseignement. *Revue suisse des sciences de l'éducation, 43*(3), 366-375.

Audran, J. (2024). Cinq enjeux d'évaluation face à l'émergence des IA génératives en éducation. *Mesure et évaluation en éducation, 47*(1), 6-26.

Béland, S., Magis, D. et Raîche, G. (2013). Estimation des paramètres d'item et de sujet à partir du modèle de Rasch : une étude comparative des logiciels BILOG-MG, ICL et R. *Mesure et évaluation en éducation, 36*(1), 83-110.

Boulet, M.-M., Lavoie, L., Labbé, P. et Lemay, F. (1990). Système informatisé pour l'évaluation formative. *Mesure et évaluation en éducation, 13*(1), 29-55.

Burguete, E., Picard, N., Andrieux, N., Fourcade, L. et Perrochon, A. (2020). Évaluation par les pairs à distance lors d'un enseignement de lecture critique d'articles pour des étudiants paramédicaux. *Évaluer _ Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 41-51.

Carrupt, R. et Barras, H. (2019). Dynamique des régulations en classe virtuelle ou en présentiel. *Évaluer - Journal international de recherche en éducation et en formation, 5*(3), 57-83.

Coen, P.-F. (1997). Analyse des régulations d'élèves scripteurs utilisant AutoéVal, logiciel d'assistance à l'écriture. *Cahiers de la recherche en éducation, 27*(3), 261-278.

Coen, P.-F. (2000). *A quoi pensent les enfants quand ils écrivent. Analyse des processus cognitifs et métacognitifs en jeu dans une tâche d'écriture assistée par le logiciel AutoéVal.* [thèse de doctorat en ligne]. Université de Fribourg. http://pfcoen.ch/a-quoi-pensent

Coen, P.-F., Detroz, P. et Younès, N. (2019). Introduction - Evaluation et numérique : des pratiques éclectiques qui explorent des espaces à déchiffrer. *Évaluer - Journal international de recherche en éducation, 5*(3), 1-8.

Coen, P.-F., Çuko, K. et Etienne-Tomasini, D. (2024). Évaluer et réguler les enseignements : l'utilisation de « OURA », un outil numérique pour apprécier les expériences d'apprentissage des apprenants.

Dans N. Loye et N. Duroisin (dir.), *Evaluation, apprentissage et numérique*, (p. 195-218). Peter Lang.

Coen, P.-F. et Pellegrini, S. (2020). Une feuille de route numérique pour évaluer formativement la progression des étudiant.e.s en contexte d'enseignement à distance. *Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 59-65.

Connes, J. (2022). *De l'IBM 360/75 au superordinateur Jean Zay.* EDP Sciences.

Da Costa Cabral, F., Gremion, C. et Roblez. A. (2020). Évaluation formatrice, rythmes et travail de groupes à distance. *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 77-84.

Dassa, C. (1988). L'intégration du diagnostic pédagogique aux apprentissages scolaires : de la théorie à la voie informatique. *Mesure et évaluation en éducation, 11*(1), 7-26.

Dassa. C. et De Cotret, S. (1994). Validation d'un système informatisé de diagnostic en mathématiques au secondaire : une approche centrée sur l'analyse didactique. *Mesure et évaluation en éducation, 16*(3-4), 5-26.

David, M. et Droyer, N. (2019). Évaluation de la co-conception d'un environnement virtuel éducatif forestier Pré-enquête à l'entrée par le critère de pertinence. *Évaluer - Journal international de recherche en éducation et en formation, 5*(3), 109-130.

De Campos, D. (1990). Outils diagnostiques et enseignement assisté par ordinateur. *Mesure et évaluation en éducation, 13*(3), 55-67.

Dechef, H. et Laveault, D. (1999). Le testing adaptatif par ordinateur. *Psychologie et psychométrie, 20*(2-3), 151-179.

Detroz, P., et Younès, N. (2023). Médiations pédagogiques en jeu à l'heure des systèmes numériques interactifs : une revue de littérature narrative et critique. *Revue français de pédagogie*, *221*, 119-155. https://doi.org/10.4000/11w0f

Dionne, E. (2006). L'implantation de l'évaluation assistée par ordinateur (EAO) au sein du système scolaire québécois : avantages et obstacles à surmonter. *Mesure et évaluation en éducation, 29*(2), 99-116.

Dionne, E. (2012). Les systèmes à réponses personnalisées (SRP) : un atout pour faire de l'évaluation formative en salle de classe ? *Mesure et évaluation en éducation, 35*(1), 47-65.

Drijvers, P. (2018). Digital assessment of mathematics: Opportunities, issues and criteria. *Mesure et évaluation en éducation, 41*(1), 41-66.

Duroisin, N. (2020). Le podcasting collaboratif, un outil pour l'évaluation formative à distance. *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 121-130.

El Hage, F. et Nahed, R. (2020). Apprentissage par Exploitation de l'Erreur et à Distance (AEED). *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 131-142.

Gardies, C., Faure, L. et Marcel, J.-F. (2019). L'usage d'un dispositif numérique synchrone pour l'évaluation de pratiques d'enseignement. *Évaluer - Journal international de recherche en éducation et en formation, 5*(3), 35-55.

Gilles, J.-L. et Charlier, B. (2020). Dispositifs d'évaluation à distance à correction automatisée versus non automatisée : analyse comparative de deux formes emblématiques. *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 143-154.

Heilporn, G. et Denis, C. (2020). Dispositifs d'évaluation en ligne : des recommandations de la littérature scientifique à des décisions précipitées en pratique. *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 165-172.

Laurier, M. (1996). Pour un diagnostic informatisé en révision de texte. *Mesure et évaluation en éducation, 18*(3), 85-106.

Lepage, I., Leduc, D. et Stockless, A. (2019). E-évaluation dynamique et engagement cognitif en contexte de grand groupe à l'université. *Évaluer - Journal international de recherche en éducation et en formation, 5*(3), 9-33.

Leroux, J. L., Desrochers, M.-E. et Myre-Bourgault, M. (2019). L'évaluation des apprentissages à l'ère du numérique en enseignement supérieur : quels besoins et quels défis ? *Évaluer - Journal international de recherche en éducation et en formation, 5*(3), 85-108.

Lison, C. (2020). La présentation orale en contexte de formation à distance : évaluer un Pecha Kucha. *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 173-180.

Loignon, G. (2022). ALSI : un nouvel outil d'analyse automatisée de la complexité linguistique pour le français québécois. *Mesure et évaluation en éducation, 44*(3), 29-57.

Martin, R. (2003). Le testing adaptatif par ordinateur dans la mesure en éducation : potentialités et limites. *Psychologie et psychométrie, 24*(2-3), 83-110.

Morin, M. et Blais, J.-G. (2017). Production semi-automatisée d'une carte conceptuelle en science et technologie. *Mesure et évaluation en éducation, 40*(1), 61-99.

Papi, C., Gérin-Lajoie, S. et Hébert, M.-H. (2020). Se rapprocher de l'évaluation à distance : dix pistes de réponse. *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 201-206.

Rached, P., Gharib, Y. et Constantin, S. (2020). Entre le cognitif et l'émotionnel... Quelle évaluation à distance en situation de crise ? *Évaluer - Journal international de recherche en éducation et en formation,* hors-série : évaluer en temps de pandémie, 223-230.

Sacré, M., Lafontaine, D. et Toczek. M.-C. (2019). Liens entre les composantes des dispositifs d'enseignement hybride et les performances des étudiants de l'enseignement supérieur : une revue systématique. *Mesure et évaluation en éducation, 42*(3), 109-152.

Sadin, E. (2015). *La vie algorithmique, critique de la raison numérique.* Éditions l'Echappée.

Schuessler, K. et Forget, J. (2009). Utilisation d'une version informatisée de la didactique de précision auprès d'élèves ayant un trouble envahissant du développement. *Mesure et évaluation en éducation, 32*(1), 25-54.

Seguin, S. P. (1984). L'utilisation des micro-ordinateurs pour l'évaluation des apprentissages : quelques perspectives. *Mesure et évaluation en éducation, 7*(1), 53-66.

Thual, O. (2020). De l'examen écrit scientifique sur table au paramétrage d'une activité Test de Moodle. *Évaluer - Journal international de*

recherche en éducation et en formation, hors-série : évaluer en temps de pandémie, 245-252.

Chapitre 22. Changer notre rapport aux preuves d'apprentissage : questions posées par les IA génératives en évaluation

Problématiser l'émergence de l'IA en évaluation des apprentissages en enseignement supérieur

Isabelle Nizet

I. Introduction

L'émergence fulgurante des intelligences artificielles génératives (IAg) au cœur de nos écosystèmes éducatifs provoque diverses réactions, notamment la crainte que les preuves d'apprentissage fournies par les personnes étudiantes ne soient plus authentiques et perdent leur validité, remettant en cause l'éthique fondatrice de l'évaluation. Une fois cet effet « répulsif » passé, une importante dynamique d'appropriation des apports et des limites des ressources génératives s'est mise en place au niveau international, national dans plusieurs pays européens et au Canada, notamment [Conseil supérieur de l'éducation et Commission de l'éthique en science et en technologie, 2024 ; Fengchun et Holmes, 2023 ; Miao et Curukova, 2024 ; Miao et Shioira, 2024]. En effet, les relations entre les IAg et l'évaluation sont prioritairement abordées sous l'angle de normes d'encadrement dans la plupart des institutions d'enseignement supérieur, essentiellement par souci éthique. De nombreux documents officiels

existent dans plusieurs pays francophones à cet égard, de même que des enquêtes et des missions nationales (par ex. en France, Pascal et al. 2025). Il en résulte le déploiement d'un large paysage de possibilités pédagogiques ayant des conséquences majeures en ce qui concerne les méthodes et les intentions d'évaluation, largement soutenu par la production de ressources éducatives libres (REL).

Du point de vue de la recherche sur cet objet, on semble en être encore à l'exploration et au recueil d'un état de l'art en progression constante, puisque peu d'expériences de grande envergure sont menées et encore moins traitées scientifiquement. Dans nos contextes francophones, la problématisation de la présence des IAg en évaluation en est à ses débuts. Des travaux récents abordent la question de manière introductive en présentant les enjeux de l'IAg face à l'évaluation à travers leur potentiel et leurs limites (Audran, 2024). En France, au Québec et en Suisse on peut relever la rare présence de thèses doctorales en cours sur le sujet spécifique de l'intégration de l'IAg à l'évaluation. Dans une optique plus large, les questions de l'usage de l'IAg par les personnes enseignantes dans leur propre pratique évaluative restent un autre sujet à explorer. Enfin, de manière encore plus large, la question des effets cognitifs et psycho-pédagogiques de l'IAg sur les trajectoires d'apprentissage reste un objet d'étude en plein développement (Merino-Campos, 2025).

Dans une perspective à la fois théorique et pragmatique, une démarche de questionnement systématique de pratiques évaluatives individuelles ou d'équipe est ici proposée à la lumière des défis que représente la présence durable et évolutive des IAg en éducation. À cet effet, nous examinerons comment les dimensions psychologiques, éthiques, axiologiques, épistémologiques et pédagogiques des preuves d'apprentissage sont remises en question. Cette démarche a été l'objet de plusieurs présentations et applications dans des conférences et des ateliers réalisés dans une dizaine d'institutions d'enseignement supérieur en France et en Suisse au cours de l'année 2024-2025, générant dans ces contextes une problématisation individuelle et collective des enjeux de l'IAg en évaluation, dans une perspective à la fois conceptuelle et pragmatique.

II. Les piliers de la validité de l'évaluation remis en question

Pour les personnes non formées en évaluation, la validité de l'évaluation au sens large repose essentiellement sur trois principes : elle doit se passer

dans un espace relativement clos dans des conditions uniformes et stables, au même moment et sur le même objet pour toutes les personnes apprenantes. Cette dimension de comparabilité est encore considérée par de nombreuses personnes enseignantes comme une condition *sine qua non* de la valeur des résultats qu'elle génère. Pour garantir cette comparabilité, le souci principal serait d'éviter des biais liés à la variabilité des conditions de passation, comme le signalent par exemple Laurier et Darria (2015). Les évaluateurs sont aussi considérés comme les seuls garants de l'expertise sur laquelle reposeraient la validité de contenu et le fonctionnement des « épreuves » (Laveault et Grégoire, 2014). D'où la tendance, parfois très prégnante dans de nombreux contextes de formation (notamment en enseignement supérieur), de restreindre les activités évaluatives à des examens écrits, dans des conditions très contrôlées, anonymisées et où la fraude par accès à des ressources externes demeure par la force de choses très limitée. Disons-le d'emblée, ce type d'évaluation n'est pas menacé par l'irruption des IAg. Cependant, la littérature scientifique reconnaît que ce type d'évaluation dite « traditionnelle » (Scallon 2015) conserve une pertinence très faible et une efficacité restreinte lorsque vient le temps d'évaluer des objets plus complexes et multidimensionnels, tels que les compétences ou le traitement de situations professionnelles plus ou moins authentiques. C'est d'ailleurs pour répondre à ce besoin d'évaluer des objets de niveaux taxonomiques supérieurs en situation que se sont mises en place progressivement et dans de nombreux programmes professionnalisants des modalités d'évaluation plus sophistiquées requérant un temps d'accomplissement beaucoup plus long (parfois plusieurs semaines) et avec un contrôle de conditions plus restreint. Au-delà des mémoires et des essais qui constituent des preuves d'apprentissage depuis de nombreuses années (Delcambre, 2018), une multitude de productions dites longues ont émergé (par ex. rapports de stages, recherches appliquées en sciences, essais sur des questions professionnelles, projets, minirecherches actions, etc.) Dans ces situations évaluatives authentiques valorisées depuis des décennies dans le cadre de l'évaluation de compétences (Guérin-Lajoie et al., 2024) et de l'évaluation durable (Boud et Soler, 2016), les personnes étudiantes produisent les preuves d'apprentissage loin des yeux de la personne enseignante, avec les outils dont ils disposent. L'évaluation critériée permet de contractualiser un lien fonctionnel entre un référent communiqué (les critères et leurs indicateurs) et un référé produit par l'apprenant (une production, une tâche complexe accomplie) (Brookart et Chen, 2015 ; Mottier Lopez et Laveault, 2008 ; Panadero et Jonsson, 2020). Dans ce contexte, la preuve d'apprentissage peut bien sûr être produite de manière frauduleuse, en tout ou en partie, aux risques et périls de l'apprenant. Cependant, des

recherches récentes menées avant l'émergence de l'IAg indiquent que les tâches authentiques pourraient réduire le besoin de tricherie (Nolla, 2021).

On peut donc affirmer que certains types de situations évaluatives sont plus vulnérables que d'autres à la tricherie, au plagiat avec ou sans IAg. D'où la mise en place systématique de contrats éthiques garantissant les conséquences punitives et éducatives de l'expérience volontaire de fraude ou de tricherie, et le développement d'un champ de recherche en évaluation sur cette question (Frenette et al. 2019 ; Rundle et al., 2023).

II.a. La dimension épistémique des preuves d'apprentissages en évaluation

Jusqu'à présent, un contrat tacite a toujours relié la personne évaluatrice et l'apprenant producteur de preuve d'apprentissage : celui d'une évaluation « juste » en réponse à une preuve « vraie » au sens où elle est bien produite par la personne qui bénéficie de l'évaluation. On pourrait dire qu'il s'agit en quelque sorte de produire de l'information ayant un caractère de preuve par l'apprenant en échange de la production d'un jugement qui aurait valeur de vérité, par l'enseignant. Il s'agirait d'une transaction à caractère épistémique, car elle produit de la connaissance par et pour la personne apprenante et la personne évaluatrice sur la preuve produite (Merle, 2018 ; Quiamzade et al., 2014). La certitude de l'identité de la personne qui produit une preuve d'apprentissage est donc au cœur du processus de légitimation du jugement de la personne évaluatrice. Or, c'est précisément cette réalité qui est menacée par l'émergence des IAg, puisque ces dernières permettent une assistance totale ou partielle dans l'accomplissement d'une tâche par réalisation robotisée et par conséquent une forme d'usurpation d'identité. On ne doit dès lors pas s'étonner que la première réaction institutionnelle ou individuelle face aux IAg, quand vient le temps de penser à leurs effets sur l'évaluation, soit de l'ordre de la fuite, de l'évitement, voire de la diabolisation, puisque le premier risque qu'elles représentent est celui de la rupture du contrat tacite qui fonde la légitimité épistémique de l'évaluation.

II.b. La dimension axiologique des preuves d'apprentissages en évaluation

Cette remise en cause du contrat épistémique est d'autant plus importante qu'elle garantit l'attribution d'un mérite ou la justification d'une sélection à l'issue de la décision évaluative.

Ainsi, dans le cadre d'un contrat évaluatif, la transaction a également une dimension axiologique, la valeur du jugement étant souvent perçue par la personne apprenante de manière dichotomique : réussir ou échouer. Cette vision dichotomique est souvent renforcée par des démarches d'interprétation normative des preuves où les étudiants d'un même groupe sont comparés entre eux, la rareté de la bonne note à l'extrême droite de la courbe normale suscitant des espoirs et des déceptions. Ainsi, cette

> rareté – au sens économique du terme – des bonnes notes contribue à créer une croyance profonde dans le système méritocratique et compétitif qui caractérise la société en général, un système dans lequel il est théoriquement possible de gagner, mais où l'échec reste un spectre qui hante tout le monde. C'est précisément dans ce contexte idéologique que la triche pourrait en effet être vue par les étudiants comme un moyen adaptatif de survivre à la pression de devoir réussir en surpassant les autres. Ce qui compte est d'être le meilleur à tout prix. (Pulfrey, 2011, p. 178-179).

Il n'est dès lors pas étonnant que la personne étudiante cherche à utiliser l'IAg de manière potentiellement frauduleuse dans ce contexte, toujours prégnant actuellement (Nizet a, à paraître).

II.c. La dimension psychopédagogique des preuves d'apprentissage en évaluation

La validité des preuves d'apprentissage repose aussi sur un contrat psychopédagogique présentant deux enjeux interdépendants : un enjeu identitaire et un enjeu social. Du point de vue identitaire, la preuve d'apprentissage produite est souvent confondue avec la personne qui la produit. L'étudiant se projette entièrement dans cette preuve qui devient littéralement « lui ». Dans cette perspective, la prise de risque de l'échec est toujours une menace psychologique pour l'estime de soi et l'identité,

comme l'évoquaient déjà Mottier Lopez et Laveault en 2008 dans leur recension de littérature. La réception d'un verdict d'échec provoque vis-à-vis de l'évaluation un sentiment de perte de contrôle et, par conséquent, un désengagement face aux tâches et à l'apprentissage (Tenret, 2011). Du point de vue social, aux yeux de la personne étudiante, l'évaluation reste une démarche arbitraire, fondée sur les privilèges de l'expertise et de l'autorité accordés à la personne enseignante, ce qui lui donne un pouvoir sur sa destinée d'apprenant. Si on ne comprend pas les sources du désir de frauder, on risque d'encourager le maintien d'une relation problématique à l'évaluation (Batruch et al., 2019 ; Butera et al., 2011]. Le développement de l'agentivité de la personne étudiante (Lepage et Colin, 2023) et d'une relation constructive aux IAg (Luo, 2024) en évaluation sont des voies à considérer pour penser une intégration des IAg au bénéfice de la personne étudiante.

Par ailleurs, dans la majorité des cas, peu de rétroaction est fournie sur la qualité de la preuve produite dans une intention d'évaluation dite « sommative » ou certificative, surtout quand la preuve produite est totalement ou partiellement inadéquate. La faible plus-value pédagogique de l'évaluation dite traditionnelle (Scallon, 2015), particulièrement parce que le processus d'interprétation reste normatif, induit aussi la tentation de mécanismes adaptatifs négatifs, tels que la tricherie. Ce rapide tour d'horizon vise essentiellement à faire comprendre que, dans de nombreuses institutions d'enseignement supérieur, la question des IAg en évaluation est posée au prisme de ce qui garantissait la validité de l'évaluation dans un système traditionnel. Notre hypothèse rejoint alors celle de nombreuses institutions qui se montrent proactives à l'égard de la question de la tolérance, voire de l'intégration des IAg en évaluation, à savoir que celles-ci deviennent incontournables et qu'il vaut mieux « domestiquer la bête que la combattre » (Lez et al., 2023).

III. Problématisation de l'introduction de l'IAg en évaluation

Nous pouvons formuler l'hypothèse que l'interdiction ou l'évitement systématique de l'IAg en évaluation représente une voie sans issue pour deux raisons : 1) l'espérance d'un contrôle total sur la production de preuves d'apprentissage s'avère extrêmement coûteuse humainement et techniquement et 2) la limitation des productions des preuves à un environnement totalement contrôlé entraîne un potentiel désalignement

pédagogique en terme de finalités de formation et de complexité des apprentissages. Ce premier type de posture à l'égard de l'IAg n'est donc pas à encourager.

Le second type de postures serait alors de considérer les IAg comme une composante inévitable de l'évaluation. Il sera alors nécessaire dans un premier temps de permettre aux différents acteurs de problématiser adéquatement comment l'introduction des IAg transforme leur relation à l'évaluation et sa pratique, pour ensuite élargir les perspectives de positionnement et d'usage raisonné de celles-ci. Dans un deuxième temps, une fois la perspective d'intégration acceptée, l'idée est d'analyser le degré de vulnérabilité des tâches évaluatives déjà utilisées par la personne enseignante pour prendre une décision quant au recours aux IAg : les éviter et ne plus y avoir recours, adapter ces tâches et les modifier pour y intégrer les IAg ou encore les remplacer par d'autres tâches non vulnérables aux IAg.

Dans chacune de ces solutions, l'idée est de rendre chaque aspect de la tâche évaluative résilient aux IAg tout en conservant sa plus-value pédagogique.

III.a. Se positionner sur un continuum d'usages des IA génératives

Actuellement, de nombreuses institutions francophones, telles que par exemple, l'Université de Sherbrooke au Québec, ont fait le choix de proposer des balises institutionnelles encadrant l'usage des IAg en enseignement (et donc en évaluation) et en recherche, de manière à soutenir le développement d'initiatives d'intégration satisfaisantes. À titre indicatif, une synthèse de ces postures peut être envisagée sous forme d'un continuum allant de l'interdiction totale vers une autorisation totalement libre. Plusieurs autres institutions ont déterminé de tels niveaux d'intégration (ou d'exclusion) à partir d'un certain nombre de paramètres dont nous proposons ici une analyse comparative.

Lors de la planification de l'évaluation certificative (ou sommative), les IAg peuvent être considérées comme une « assistance » pour réaliser les tâches d'évaluation. À ce titre, elles peuvent être « évitées », « contournées » ou « intégrées » (Vangrunderbeek et al. 2024). L'évitement consistera à créer des conditions suffisamment fiables pour rendre l'accès aux IAg impossible durant la production des preuves d'apprentissage

(Vangrunderbeek et al. 2024). Plusieurs institutions parleront alors d'interdiction pour signifier que tout usage des IAg est considéré comme frauduleux, et donc potentiellement délictueux dans ces situations.

En réponse à cette interdiction, le retour systématique aux examens écrits en lieux fermés et contrôlés et la passation d'examens oraux en présence sont généralement adoptés comme solution. Interdire l'IAg pourrait alors constituer une réponse qui paraît logique, reflétant d'abord un souci éthique légitime de préserver intactes les conditions qui ont garanti la valeur de l'évaluation jusqu'à présent. Ces pratiques peuvent cependant soulever un certain nombre de questions relatives à l'alignement pédagogique. Les examens écrits ne risquent-ils pas de limiter l'évaluation à des objets atomisés ou de niveau taxonomique bas ? Par exemple, en limitant les questions d'évaluation à des questions fermées portant sur des définitions de concepts ou l'explication de phénomènes simples. Des pans entiers d'apprentissage ne pourront plus être prouvés par les apprenants, notamment la mobilisation interconnectée de connaissances complexes, qu'elles soient déclaratives ou procédurales. L'enjeu est encore plus important pour l'activation de connaissances conditionnelles, qui fondent l'expertise et ne peuvent être mobilisées que dans des situations présentant des défis d'analyse, de jugement et de création, particulièrement dans des curricula par compétences. Les questions des examens oraux étant distribuées arbitrairement, comment sera-t-il possible d'évaluer des dimensions essentielles des compétences de manière valide (Wiertz et al., 2022) ? Il n'est par ailleurs pas toujours possible, ni même souhaitable, de placer tous les apprenants en situation réelle et individuelle d'évaluation, ni même pertinent de recueillir des traces verbales sur certains apprentissages. Enfin, le caractère simultané du recueil d'information et de son interprétation pose de sérieux problèmes de biais et de traitement d'information de manière transparente, équitable et juste (Chaumont et Leroux, 2018). Nous pouvons donc provisoirement conclure à ce stade du développement du propos que l'interdiction pure et simple de l'IAg en évaluation pourrait représenter une potentielle « régression pédagogique » par simplification des évaluations liées aux conditions de contrôle du plagiat par l'IAg.

Le *contournement* consiste à rendre les tâches d'évaluation suffisamment complexes et contextualisées pour que les IA ne puissent pas produire de réponses de qualité, le recours à l'IAg devenant inutile, voire impossible (Vangrunderbeek et al. 2024). Il semble cependant illusoire de croire que nous pourrons créer des tâches d'évaluation si complexes qu'elles échappent aux progrès constants de l'IAg (Amatrian, 2024).

Dans plusieurs institutions, on parlera davantage d'un *usage limité et encadré* par des consignes qui s'adressent à la personne étudiante (Cabana et Côté, 2024). Ainsi, le fardeau de l'adaptation des tâches évaluatives est distribué à la fois sur la personne enseignante qui rend ces tâches résilientes à l'IAg tout en impliquant la personne étudiante dans un apprentissage de l'usage des IAg. Par conséquent, ces institutions insistent davantage sur une *intégration balisée* (Cabana et Côté, 2024). Par exemple, l'intégration consiste à adapter les activités d'apprentissage et d'évaluation en favorisant un usage critique et responsable des IAg (Vangrunderbeek et al. 2024). Ce troisième niveau d'usage donne alors lieu à une déclinaison plus diversifiée dans certaines institutions : une utilisation limitée pour assister l'apprentissage ; une utilisation guidée pour améliorer a posteriori une preuve d'apprentissage avec déclaration d'usage obligatoire, par le biais de rétroactions fournies par l'IA sur un travail, la correction de texte, etc. ; une utilisation « balisée » dans laquelle l'IA est utilisée *a priori* pour produire une preuve d'apprentissage qui sera ensuite améliorée par la personne étudiante, par exemple pour de la recherche, la production de résumés ou l'exploration d'idées. L'Université de Sherbrooke parle de génération de textes, d'images, de résolution de problème (Cabana et Coté, 2024). Les traces de production issues des IAg doivent alors être obligatoirement recueillies et transmises ; enfin, au dernier stade : une utilisation libre des IAg dans la production de preuves avec déclaration obligatoire de l'usage et des traces recueillies.

Il nous semble important de remarquer que ces catégorisations d'usages se recoupent et que les exemples qui illustrent ces usages ne reposent pas sur une compréhension univoque de ce que sont les tâches et les usages eux-mêmes. Ces catégorisations tentent de décrire les caractéristiques d'usage des IAg à partir de quatre critères : 1) la nature taxonomique des tâches (résumer, générer, évaluer, planifier, etc.) ; 2) la fonction des IA dans le processus de production des preuves (avant, pendant, après ; assistance partielle, totale) ; 3) le rôle de la personne étudiante qui sollicite les IAg ou les utilise dans son processus de production (améliorer, diriger, modifier, etc.) et 4) la documentation de l'usage avec déclaration et recueil de traces ou preuves que nous appellerons « secondaires » ; ce dernier paramètre englobe la dimension éthique de l'usage, parfois encadrée par des spécifications éditoriales (Peters, 2023).

Notre hypothèse est que les balises ainsi formulées permettent de sécuriser en quelque sorte une réflexion sur l'usage des IAg pour les personnes étudiantes et enseignantes qui entament une réflexion collective ou individuelle sur le sujet. Mais la réflexion individuelle gagnerait

davantage à être soutenue par un questionnement systématique sur les modalités d'évaluation déjà utilisées.

Afin de faciliter la prise de décision des personnes enseignantes et des responsables institutionnels, nous proposons ici une représentation simplifiée de cette utilisation progressive sous forme de synthèse.

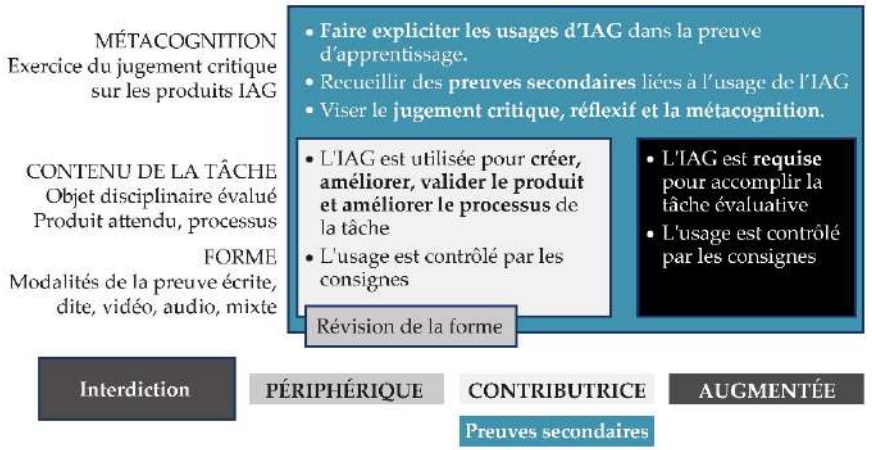


Figure 1 : Intégration d'IA génératives aux preuves d'apprentissages en fonction des objets d'évaluation.

Tout d'abord, il semble y avoir consensus sur le fait de maintenir la possibilité d'une **interdiction explicite** des IAg. À partir du moment où l'on ne souhaite pas se restreindre au choix d'interdire le recours à L'IA en évaluation, plusieurs pistes s'ouvrent à la personne enseignante. L'usage des IAg peut être autorisé, voire encouragé pour améliorer la **forme** des preuves en fonction de leurs modalités de production. Nous pensons que les IAg peuvent être autorisées pour des usages indépendants des apprentissages disciplinaires, ou autrement dit « transversaux » aux disciplines parce qu'elles peuvent assister la personne étudiante dans des démarches de révision de la qualité de la langue dans les travaux écrits, par exemple, ou encore dans la planification d'un travail long, l'organisation d'une recherche, la vérification de référencement, etc. Nous parlerions alors **d'usage périphérique** réservé à améliorer la forme des preuves. Cela signifie que les requêtes créées ne nécessitent pas le recours à des contenus disciplinaires particuliers. Autrement dit : le recours aux IAg ne repose pas sur la mobilisation de connaissances propres à une

discipline, mais plutôt sur des connaissances liées au métier d'étudiant comme tel.

De manière plus ciblée, l'accès aux IA pourrait viser leur **contribution explicite** aux processus et à la production de la preuve d'apprentissage d'un point de vue disciplinaire. Les IAg sont alors utilisées de manière proactive ou rétroactive dans le processus de production de la preuve pour créer, améliorer, valider celle-ci en tout ou en partie, car les différentes étapes d'un processus dans une tâche évaluative peuvent (ou non) bénéficier de leur contribution. C'est à la personne enseignante de le déterminer dans la construction de ses consignes. Ces améliorations et validations requièrent de la part de la personne étudiante la création de requêtes rédigées en mobilisant des concepts, des principes, des règles, des critères propres à la discipline enseignée. On peut donc considérer que ce matériau de dialogue avec les IAg constitue en soi un ensemble d'informations qui peuvent être recueillies pour poser un jugement évaluatif sur l'apprentissage de connaissances spécifiques à la discipline. C'est pourquoi, à ce rôle contributif, pourra être associé à l'usage des IAg une fonction de preuves secondaires d'apprentissage. C'est à ce niveau que les consignes devraient explicitement mentionner la nécessité de réaliser une démarche critique et métacognitive autonome de la part de la personne étudiante, ce qui représente la complexification des tâches évoquée plus haut. Notons que, jusqu'à ce niveau d'intégration, les IAg ne font pas encore — comme telles — partie d'un enseignement explicite. Il s'agit toujours d'autoriser un usage qui demeure circonscrit à une littératie de base (Haugsbaken et Hagelia, 2024 ; Lepage et Roy, 2024). Enfin, au dernier niveau d'intégration, on observe que les formations professionnalisantes intègrent dans leur curriculum des apprentissages d'IAg spécifiques aux disciplines et aux professions, par exemple lorsque, en formation au design ou en architecture, des IA génératives de production d'images ou de plans sont enseignées en formation puisque présentes dans le contexte professionnel. Et il sera alors capital d'intégrer dans les tâches évaluatives les apprentissages spécifiques à ces IAg comme finalités explicites de formation. C'est dans cette perspective que nous parlerions d'une **évaluation « augmentée »**.

III.b. Identifier le niveau de vulnérabilité des tâches évaluatives à l'IAG

L'étape de l'élaboration des tâches évaluatives revêt un caractère essentiel dans la démarche de positionnement de la personne enseignante face aux

IAg. Le degré de complexité (de la reproduction à la production) des tâches et sous-tâches à accomplir, est repérable dans le niveau taxonomique des verbes d'action qui servent à les décrire. Dans le cas de production de preuves multidimensionnelles, l'évaluation dite authentique porte sur des objets complexes, professionnellement significatifs, engageant l'apprenant dans le traitement de situations réelles ou réalistes qui demandent du temps, de l'espace, dans des conditions difficilement contrôlables par la personne enseignante. Les situations à traiter sont souvent très spécifiquement contextualisées et peuvent avoir un caractère unique, tandis que d'autres sont structurées selon des paramètres comparables (notamment dans le cas de simulations).

Il s'agit aussi des modalités de recueil de preuves d'apprentissage les moins surveillées et qui ont le plus haut niveau de plus-value pédagogique dans un contexte professionnalisant. Par exemple : les essais, les dissertations, les rapports de tout type (laboratoire, stages, terrain, etc.), les études de cas, toutes les situations de création, d'analyse avec production écrite en preuve, les QCM réalisés en ligne hors classe, toutes les tâches avec preuves orales exigeant une préparation séparée, voire les tâches réflexives, les journaux de bord et les autoévaluations. En somme, cela nous amène à constater que les récents progrès dans le développement d'une évaluation plus significative, engageante et authentique requise par la complexité sans cesse croissante des référentiels de formation sont directement menacés par l'irruption des IAg. Mauvaise nouvelle. D'autant plus mauvaise que, dans plusieurs milieux institutionnels, on peine déjà à s'émanciper d'une évaluation traditionnelle qui se borne à évaluer des connaissances à des niveaux taxonomiques bas, soit par ignorance de modalités plus adéquates, faute de véritable formation en évaluation des apprentissages, soit parce que la grande taille des cohortes d'étudiants rend fastidieuse la « correction » de travaux plus longs et plus complexes. Les évaluations en présentiel et surveillées, orales, extrêmement personnalisées, évaluées avec des instruments de détection et explicitement encadrées par des normes éthiques sont quant à elles considérées comme faiblement, voire pas du tout vulnérables, ce qui explique l'attraction dont elles sont l'objet.

Le niveau d'authenticité et de complexité d'une situation évaluative combiné au niveau de contrôle dans sa réalisation déterminent donc le niveau de vulnérabilité de cette situation aux IAg.

III.c. Conserver la plus-value pédagogique des tâches évaluatives

Lorsque le traitement de ces situations est utilisé comme preuve d'apprentissage dans des traces écrites, verbales ou motrices, une **interprétation** critériée est généralement utilisée pour restreindre les objets de jugement à des dimensions observables et communicables. L'interprétation des traces repose sur des décisions pédagogiques en matière de degré de transparence des référents : vais-je ou ne vais-je pas communiquer les critères d'évaluation et de bons exemples aux personnes étudiantes avant qu'ils ne produisent la preuve d'apprentissage ? Vais-je ou non partager l'exercice du jugement avec elles selon des modalités de coévaluation, par exemple ? L'exercice du jugement sera-t-il soutenu par des instruments comme des grilles descriptives ? Sous quelles formes sera exprimé le jugement évaluatif ?

C'est notamment cette perspective critériée qui donne aux tâches évaluatives authentiques et complexes la plus-value pédagogique recherchée. En effet, le contrat évaluatif y est généralement matérialisé dans la communication transparente de critères, d'indicateurs et d'échelles de performance explicites dans des artefacts partagés. Ce type d'instrument a plusieurs effets positifs sur les personnes étudiantes : la transparence réduit l'insécurité face aux attentes, ce qui réduit le stress. La personne étudiante développe une perception que l'évaluation n'est pas arbitraire, ce qui rend la compensation par la tricherie inutile. Elle développe un sentiment de capacité de régulation et de contrôle sur la réussite grâce aux éléments descriptifs de la grille. Le processus de réalisation de la tâche est valorisé et pas seulement le résultat, ce qui nécessite d'être agent du processus par la production de traces relatives à celui-ci. Enfin, lorsque les critères sont coconstruits ou discutés au préalable, un sentiment de relation à des référents communs se développe (Frenette et al., 2019).

De plus, à l'issue du jugement évaluatif, des décisions pédagogiques sont généralement prises, par exemple : quelle régulation sera possible après l'évaluation, quelle forme prendront les résultats (notation ou rétroaction), la décision de réussite ou d'échec est-elle prise en fonction d'un seuil de passage, d'une échelle descriptive et les résultats seront-ils répartis selon une courbe normale ou une courbe en J ? Ces décisions influencent l'existence ou l'absence de la plus-value pédagogique de l'évaluation.

Au terme de cette problématisation, nous pouvons donc retenir que les décisions relatives à l'exclusion ou à l'intégration des IAg dans l'évaluation des apprentissages requièrent trois étapes :

1. Se positionner face à l'IA générative dans sa discipline et les cours offerts en fonction d'un cadre institutionnel non contraignant, mais structurant ;

2. Identifier le niveau de vulnérabilité des tâches évaluatives déjà en place ;

3. En fonction de ce diagnostic, prendre des décisions pour conserver la plus-value pédagogique des tâches évaluatives dans un contexte où l'IAg y est explicitement utilisée et contrôlée.

Dans une perspective pragmatique, la nécessité est d'adopter une démarche qui permettra de rendre les tâches évaluatives résilientes aux IAg plutôt que de les exclure d'emblée. Pour clôturer ce chapitre, nous proposons donc une démarche pratique qui pourra être utilisée dans différents contextes.

L'armature d'une situation d'évaluation étant constituée de tâches, de consignes et de preuves produites (incluant des processus et des productions), toutes ces dimensions sont concernées par l'interrogation sur l'IAg.

Questions à se poser	Conditions
L'usage des IAg est-il interdit ?	Rédiger des consignes pour étapes et produits sans IAg.
L'IAg pourrait-elle être autorisée pour réviser la forme finale ?	Rédiger des consignes et vérifier que les personnes étudiantes ont accès aux outils librement et gratuitement.
L'IAG peut-elle contribuer aux produits intermédiaires ou au produit final ?	Identifier les étapes et produits intermédiaires dans la (les) tâche(s).
Pour chaque produit et chaque étape intermédiaire du processus.	Identifier la vulnérabilité de chaque produit et chaque étape du processus aux IAg en testant les consignes.
Pour chaque étape et chaque produit vulnérable : décider du maintien ou de la suppression.	Supprimer le produit ou l'étape en cas de vulnérabilité haute, remplacer ou modifier par des consignes incluant explicitement le recours à l'IAg.
Pour chaque étape du processus ou produit : identifier les connaissances que l'étudiant va devoir mobiliser pour créer des requêtes pertinentes.	Rédiger des consignes pour étapes et produits générables par les IAg. Identifier les traces secondaires à recueillir. Préciser des critères pour interpréter les traces recueillies.
Idem : quelles connaissances l'étudiant va-t-il devoir mobiliser pour dialoguer avec l'IAg ?	Rédiger des consignes pour encadrer la production de requêtes (prompts). Identifier les traces secondaires à recueillir. Préciser des critères pour interpréter les traces recueillies.
Quelles connaissances mobiliser pour analyser, évaluer et améliorer les produits et processus générés par l'IAg ?	Rédiger des consignes pour encadrer l'analyse du produit des requêtes. Identifier les traces secondaires à recueillir Préciser des critères pour interpréter les traces recueillies.
L'usage de l'IAG doit-il être évalué ?	Rédiger des consignes visant l'usage des IAg comme objet d'apprentissage.

Tableau 1 : Démarche d'intégration de l'IA générative en évaluation : questionnement pédagogique sur les tâches, les consignes et les preuves à produire

IV. Conclusion

La problématisation des enjeux de l'irruption des IAg en évaluation des apprentissages est largement centrée sur ses effets frauduleux. Nous avons

voulu montrer que les recherches portant sur la fraude demeurent pertinentes pour comprendre en quoi l'usage des IAg traduit, comme les autres formes de tricherie, une forme d'adaptation négative de la part de la personne étudiante. Nous avons aussi abordé les limites d'une conception non instruite de la validité pour induire une ouverture à un usage raisonné des IAg en évaluation. Conserver les acquis de l'évaluation authentique en maintenant la plus-value de celle-ci et en la rendant résiliente aux IAg est une alternative à une perspective qui réduirait la portée taxonomique et pédagogique de l'évaluation. Ainsi, dans la perspective de soutenir le développement d'un raisonnement structuré pour réfléchir à l'intégration des IAg en évaluation dans la formation en enseignement supérieur, plusieurs défis se présentent. Pour les personnes enseignantes : 1) la nécessité de la maîtrise d'outils d'IAg afin d'en proposer proactivement et intentionnellement un usage spécifique en évaluation ; 2) une incitation à renforcer l'alignement pédagogique de l'évaluation avec l'enseignement et l'apprentissage, y compris celui des IAg ; 3) la nécessité de planifier l'évaluation à des niveaux taxonomiques variés, et pas nécessairement plus élevés, étant donné la capacité d'automatisation de tâches généralement évaluée en enseignement supérieur. D'autre part, pour les personnes étudiantes : 1) une modification importante de leur relation à l'évaluation et aux IAg dans la mesure où il ne s'agirait plus d'essayer de prouver une réussite en recourant à des moyens non légitimes ; 2) une incitation à mieux comprendre la valeur de ces moyens au regard de leur processus d'apprentissage et des preuves à produire.

V. Bibliographie

Amatrian, X. (2024). Prompt Design and Engineering: Introduction and Advanced Methods. *ArXiv, abs/2401.14423.*

Audran, J. (2024). Cinq enjeux d'évaluation face à l'émergence des IA génératives en éducation. *Mesure et évaluation en éducation*, *47*(1), 6–26. https://doi.org/10.7202/1114564ar

Batruch, A., Autin, F., Bataillard, F. et Butera, F. (2019). School selection and the social class divide: How tracking contributes to the reproduction of inequalities. *Personality and Social Psychology Bulletin*, *45*, 477–490.

Boud, D. et Soler, R. (2016) Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413, http://dx.doi.org/10.1080/02602938.2015.1018133

Brookhart, S. M. et Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review, 67*(3), 343-368. https://doi.org/10.1080/00131911.2014.929565

Butera, F., Buchs, C. et Darnon, C. (dir.) (2011). *L'évaluation, une menace ?* Presses Universitaires de France.

Cabana, M. et Côté, J.-A. (2024). *Balises d'utilisation des outils d'intelligence artificielle générative*. Service de soutien à la formation, Université de Sherbrooke.

Chaumont, M. et Leroux, J.-L. (2018) Le jugement évaluatif : subjectivité, biais cognitifs et postures du professeur. *Pédagogie collégiale 31*(3), 27-33).

Delcambre, I. (2018). Littéracies universitaires et mémoires professionnels, *Questions Vives* (30) https://doi.org/10.4000/questionsvives.3072

Frenette, E., Fontaine, S., Hébert, M.-H. et Éthier, M. (2019). Étude sur la propension à tricher aux examens à l'université : élaboration et processus de validation du Questionnaire sur la tricherie aux examens à l'université (QTEU). *Mesure et évaluation en éducation, 42*(2), 1–33. https://doi.org/10.7202/1071514ar

Guérin-Lajoie, S., Hébert, M.-H et Papi, C. (2024). L'évaluation en situation authentique. *Le Tableau, échange de bonnes pratiques entre enseignants de niveau universitaire 13*(1), Direction du soutien aux études et des bibliothèques (DSEB) en collaboration avec le Groupe d'intervention et d'innovation pédagogique (GRIIP). https://pedagogie.uquebec.ca/sites/default/files/documents/numeros-tableau/Le_Tableau_Vol13_no1_EvaluationAuthentiqueVF_0.pdf

Haugsbaken, H. et Hagelia, M. (2024). *A New AI Literacy for the Algorithmic Age: Prompt Engineering or Educational Promptization*? 4 th International Conference on Applied Artificial Intelligence (ICAPAI), 1-8. https://doi.org/10.1109/ICAPAI61893.2024.10541229

Kaboré, N. D., Frenette, É. et Hébert, M.-H. (2022). Revue systématique sur les pratiques évaluatives probantes en évaluation formative à l'enseignement primaire et secondaire. *Revue des sciences de l'éducation*, *48*(1). https://doi.org/10.7202/1096356ar

Laurier, M. et Diarra, L. (2015). Évaluer l'écriture à l'ordinateur dans les épreuves uniformes : promesses et défis. *Québec français, (175)*,5–17.

Laveault, D. et Grégoire, J. (2014). *Introduction aux théories des tests en psychologie et en sciences de l'éducation.* De Boeck.

Leblanc, F. (2024). *Encadrer l'utilisation de ChatGPT et autres IAG.* Service de soutien à la formation. Université de Sherbrooke.

Lepage, A. et Roy, N. (2024). Le développement d'une échelle de mesure de la littératie de l'intelligence artificielle chez les enseignants et les enseignantes du postsecondaire. *Mesure et évaluation en éducation, 47*(2), 39–69. https://doi.org/10.7202/1117467ar

Lepage, A. et Collin, S. (2023). Préserver l'agentivité des enseignants et des élèves : des pistes issues d'une recension des écrits. Dans M. Romero et L. Heiser (dir.), *Enseigner et apprendre à l'ère de l'intelligence artificielle*. Canopé, Livre blanc. hal-04013223v2 p. 81-96.

Lez, A., Dubé, E. et Beaulieu, M. (2023). *Évaluer à l'ère des IA : un guide de réflexion*. Service de soutien à la formation, Université de Sherbrooke.

Luo (Jess), J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, *30*(4), 991–1006. https://doi.org/10.1080/13562517.2024.2341005

Merle, P. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives*. Presses Universitaires de France.

Merino-Campos, C. (2025). The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review. *Trends in Higher Education*, *4*(2), 17. https://doi.org/10.3390/higheredu4020017

Miao, F. et Cukurova, M. (2024). *Référentiel de compétences en IA pour les enseignants*. UNESCO. https://doi.org/10.54675/BQZD8407

Miao, F., Shiohira, K. et Lao, N. (2024). *Référentiel de compétences en IA pour les apprenants*. UNESCO. https://doi.org/10.54675/NXRY6511

Mottier Lopez, L. et Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et évaluation en éducation, 31*(3), 5–34. https://doi.org/10.7202/1024962ar

Nizet, I. (à paraître a). L'évaluation aujourd'hui et demain. Quelles ruptures contextuelles ? Quelles ruptures culturelles ? Dans E. Issaieva, I. Nizet et L. Mottier Lopez (dir.), *Pluralités de contextes, pluralités d'évaluation en éducation : quelles interactions et quels enjeux ?* Presses de l'ADMEE.

Nizet, I. (à paraître b). Évaluer en recherche : la production de connaissances dans un contexte post quantophrénique. Dans R. Malet et J.-F. Marcel (dir.), *Les nouveaux défis de l'évaluation. Recherche en éducation, résultats et politiques des savoirs.* Cépaduès Édition.

Nolla, J.- M. (2021). La lutte contre le plagiat étudiant dans l'évaluation. Dans F. Lafleur, J.-M. Nolla et G. Samson (dir.), *Évaluation des apprentissages en formation à distance : Enjeux, modalités et opportunités de formation en enseignement supérieur* (p. 57-73). Presses de l'Université du Québec.

Panadero, E. et Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, *30*. https://doi.org/10.1016/j.edurev.2020.100329

Pascal, F., Taddei, F., de Falco, M. et Gallié, É.-P. (2025). *IA et Enseignement Supérieur : Formation, Structuration et Appropriation par la Société*. Rapport du Ministère chargé de l'enseignement supérieur et de la recherche.

Peters, M. (2023). *Logos pour une utilisation transparente de l'Intelligence artificielle.* Université d'Ottawa. https://mpeters.uqo.ca/utilisation-transparente-de-lintelligence-artificielle/

Pulfrey, C., Buchs, C. et Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology, 103*(3), 683–700. https://doi.org/10.1037/a0023911

Pulfrey, C. (2011). Promotion de soi et triche. Dans F. Butera, C. Buchs et C. Darnon (dir.), *L'évaluation, une menace ?* (p. 177-186). Presses Universitaires de France.

Quiamzade, A., Mugny, G. et Butera, F. (2014). *Psychologie sociale de la connaissance : Étayage expérimental*. Presses universitaires de Grenoble. https://doi.org/10.3917/pug.quiam.2014.01 .

Rundle, K., Curtis, G. J. et Clare, J. (2023). Why students do not engage in contract cheating: A closer look. *International Journal for Educational Integrity*, *19*(1), 11. https://doi.org/10.1007/s40979-023-00132-5

Scallon, G. (2015). *Des savoirs aux compétences. Explorations en évaluation des apprentissages.* Éditions du Renouveau Pédagogique Inc. (ERPI).

Tenret, E. (2011). *Les étudiants et le mérite. A quoi bon être diplômé ? La Documentation française, coll. « études et recherches — OVE »*, 1 page. EAN : 9782110081988.

Vangrunderbeek, P. (2024). *Les clés du LLL : Intégrer l'IA générative dans les stratégies pédagogiques.* Louvain Learning Lab. UCLOuvain.

Wiertz, C., Blondeau, B., Francotte, E., Galand, B. et Colognesi, S. (2022). Utiliser une grille critériée pour évaluer les explications orales de ses pairs : Quels fonctionnements et quels effets ? *e-JIREF, 8*(2), Art. 2.

Chapitre 23. 40 ans de crises de valeurs en évaluation

Crisologie de nos recherches

Alban Roblez

Avertissement : Ce chapitre part de nos réflexions présentées lors des cycles de conférences organisées par la section Suisse de l'ADMEE-Europe en 2023-2024[1] et de notre article « *L'évaluation comme une activité illustratrice de crises : une place pour l'inanticipable ?* » (Roblez, 2024c).

I. Introduction

L'évaluation fait partie de ces activités qui font parler, voire qui sont controversantes, et ce, dès leur genèse. Que ce soit dans le début des années 1900 lors des vagues de testing au sein des écoles aux Etats-Unis d'Amérique du Nord afin de remplacer le jugement professoral par un outil (Pelletier, 1971) jusqu'aux évaluations nationales et internationales standardisées des élèves ; que ce soit dans les processus de sélection à l'entrée en formation dans les collèges jésuites pour une aristocratie religieuse (Merle, 2018) jusqu'aux algorithmes de répartition aux entrées à l'université française (Tenret, 2021) ; que ce soit par les instances d'évaluation par les pairs pour évaluer la qualité et l'éthique des projets de

[1] Notre conférence était intitulée « C'est la crise : Un contexte de valeurs en tension dans l'évaluation » (Roblez, 2024b).

recherche (Allaire, 2022) jusqu'aux institutions de contrôle de la qualité de la recherche elle-même (Roblez, 2024a) : l'évaluation, quelles que soient ses formes, ne laisse pas indifférent… ni ne repose sur des assises aussi stables qu'il y paraît. Tant et si bien que l'évaluation pourrait être envisagée comme une activité sociale singulière permettant, à rebours des critiques dont elle est sujette et des multiples *disputatio*, de rendre compte des états de crise : ceux qu'elle génère bien sûr, mais aussi ceux plus sociétaux dans lesquels elle gravite.

Et si l'évaluation était une activité permettant de mieux comprendre l'état des crises de nos sociétés ? L'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE) Europe et sa grande sœur l'ADMEE-Canada sont en cela un formidable laboratoire sociohistorique de ces différentes mouvances « de crises ».

Mais de quelles crises parlons-nous ? Et que serait une « crise en évaluation » : celle de ses théories ? de ses pratiques ? de ses effets sociaux et psycho-sociaux ? Qu'est-ce que cela vient dire de ces quarante années de recherches en évaluation dans les paysages professionnels couverts par les membres des ADMEE ? En route pour une clinique des crises en évaluation, pour un voyage en « crisologie »[2] (Morin, 2012) évaluative !

II. Crise et critique

Le mot comme le concept de crise proviennent tous deux des recherches en biologie (Bernard de Raymond, 2023) afin d'illustrer des processus évolutifs au sein d'un organisme. Ainsi, parler d'une crise n'a de sens que lorsque le concept est utilisé pour saisir une compréhension d'un état se transformant dans un environnement lui-même sujet au mouvement[3]. À partir d'une sélection de littératures scientifiques pluridisciplinaires, la crise pourrait se définir à la fois comme une « situation critique », qu'elle le soit par son état réunissant des antagonismes en son sein (des critiques qui affectent l'ordre ou la stabilité de ce qui est critiqué et des personnes qui y croient) ou par sa nature même de révéler ses structures ou ses

[2] C'est-à-dire tout autant une méthode pour aborder les crises qu'une caractériologie de celles-ci, proposée par Morin (1976).

[3] C'est la raison pour laquelle Morin insiste pour ancrer le concept dans un environnement vivant et donc sujet aux incertitudes. La crise n'est pas le mot du débordement, de l'échappé : elle intègre cela, et bien plus encore (Morin, 2012).

fondements : parler d'une crise de l'éducation revient à considérer que l'éducation, telle qu'elle est perçue par ses critiques, ne convient pas ou plus et, ce faisant, comporte une vision, une considération de ce qu'est l'éducation pour les parties prenantes.

Par conséquent, une crise se situe dans le temps et dans l'espace. Elle reçoit plusieurs acceptions. Heurtaux et al. (2023) en identifient trois principales, liées à différents courants épistémologiques. Le premier courant met la crise comme un moment proprement historique. Ce faisant, elle est repérable dans le cours de l'Histoire : on peut l'identifier, trouver son ou ses lieux d'effets, les personnes concernées. On y trouve aussi son point de rupture : ce qu'on peut dénommer le « Turning Point » (Abbott, 2009). C'est l'instant de rupture franchie, la tension est atteinte à son acmé : il y aura un avant et un après. Les travaux de Kosseleck (2006) et de Ricœur (2019) incarnent bien cette acception.

Le deuxième courant s'assimile à un jugement social, voire global, d'une « fin du monde », sinon fin « d'un » monde. Proche des théories de l'effondrement (Diamond, 2009 ; Servigne et Stevens, 2015), cette approche de la crise la positionne comme un passage forcé, pouvant être perçu et éprouvé comme violent. Une crise se pose en cycle de crises : elles s'entraînent les unes les autres, par « contagion » (Bernard de Raymond, 2023, p. 13). La crise est à la fois une transition d'un état d'un monde vers un autre, comme l'état du monde concerné comme limité ou ayant atteint ses limites.

Un dernier courant, héritier des philosophies modernes[4], pose la crise comme une situation épistémique critique. Les connaissances et les savoirs cumulés et utilisés par les individus ne fonctionnent plus tout à fait, ou plus du tout, pour comprendre et expliquer des faits. Ce dysfonctionnement peut provenir d'un sursaut critique sur les origines ou les fondements des connaissances, comme provenant d'une personne dont la morale ou l'éthique se sont avérées négatives, condamnables ; ou dont les constitutions se sont opérées en contexte colonial, en domination ou en occultation d'une partie de la population.

Le point commun de ces trois acceptions ou approches du concept de crise est sa nature sociale et axiologique. Une crise génère des discours, des

[4] Notamment la phénoménologie européenne de Husserl (1976) puis Arendt (2013), de l'existentialisme avec Sartre (1970) et de la philosophie des sciences notamment avec Popper (1998) et Kuhn (2018) ou de la philosophie critique avec Foucault (2008).

positionnements idéologiques ou politiques, de façon plus générale : elle génère des valeurs sur elle et des valeurs sont fragilisées ou cristallisées par elle, dans le vécu des personnes. Partant, deux choses nous paraissent structurelles d'une crise : il y a toujours des personnes concernées par elle, qui chercheront à la (dé)légitimer (Fassin, 2023) ; il y a une mise en visibilité d'informations, de faits, de situations dans une crise (appelant ainsi à être attentif au « reste » plus implicite : ce qui est caché, ce qui est oublié, ce qui est critiqué. « Comme le rappelle Anatole Bailly, en grec ancien, le substantif krisis signifie à la fois l'action de distinguer, de séparer, et l'action de décider, de juger. […] Se combinent donc un sens analytique et un sens normatif » (Fassin, 2023, p. 19 -20). Tout comme l'évaluation ?

III. La crise (en recherche) dans l'évaluation

À ce rappel étymologique de Fassin sur la notion de crise, comment ne pas penser à l'évaluation ? « Distinguer », « séparer » notamment les informations valides et invalides, veiller à identifier les circonstances favorisant une démarche authentique de la performance chez la personne dite évaluée, ou alors distinguer son jugement personnel d'un jugement professionnel encadré par une méthode, un instrument scientifique… ceci est connu par des membres de l'ADMEE ! « Décider », « juger » mais aussi réguler fondent aussi le geste central de l'évaluation : pour encourager à apprendre durablement et à (s') améliorer, pour orienter la personne ou ses documentations, pour certifier un résultat, une procédure. Les deux « sens analytique et […] normatif » que Fassin identifie (2023, p. 20) se retrouvent dans les « démarches » et « fonctions » de l'évaluation (de Ketele, 2010, p. 30).

Le concept lui-même n'est pas inconnu de nos travaux de recherche : nous référençons cinq publications[5] attachées à l'ADMEE exposant la crise (ou une de ses manifestations) comme l'objet étudié. Mais c'est sur deux d'entre elles que nous nous attarderons, puisqu'elles balisaient l'époque de l'ADMEE-Europe : l'article puis l'ouvrage dirigé par Figari et Achouche pour les 10 ans de l'Association (1997, 2001) ; un peu plus de dix années plus tard, l'ouvrage collectif dirigé par Mottier Lopez et Figari (2012b) où

[5] Présentées en fin de ce chapitre.

les deux collègues posaient l'état des questionnements épistémologiques d'alors.

Au lendemain du dixième anniversaire de l'ADMEE-Europe (et de son colloque), Figari et Achouche rendaient compte des changements et des interrogations portés dans l'évaluation (Figari et Achouche, 1997, 2001). Ces derniers se développaient sous deux angles : au sein des contextes scolaires et dans les contextes socioprofessionnels. L'évaluation était observée comme se complexifiant (Ardoino, 2001), à la suite de son « élargissement » repéré par Cardinet (Laveault, 2014 ; Mottier Lopez et al., 2017). Dans la première conclusion, Ardoino soulevait le « tournant critique » (2001, p. 117) sur les niveaux d'analyse et des zones d'actions de l'évaluation en contextes éducatifs, démontrant un « passage de la méthodologie de l'évaluation […] à celui d'interrogations plus résolument épistémologiques » (2001, p. 117). Dans la seconde conclusion, Figari soulignait la nature transactionnelle de l'évaluation, générative « d'effets sociaux » (Figari, 2001, p. 252 à partir des contributions de Barbier et Demailly dans le même ouvrage) ainsi que deux « types de questionnement pour la recherche » : ceux portant sur les « mécanismes d'attribution de valeur (pour Lecointe) » et ceux sur « l'économisation de la valeur dans les pratiques d'évaluation » chez Triby (2001, p. 252). S'interroge alors la jonction entre les pratiques de recherche et les pratiques d'évaluation, toutes deux génératrices de connaissances et impliquant des enjeux éthiques, méthodologiques et politiques.

En 2012 Mottier Lopez et Figari éditaient et dirigeaient un ouvrage collectif restituant là aussi l'état des questionnements en matière d'épistémologie. Leur chapitre introductif s'intitule « Modèles et modélisations face à une crise du sens de l'évaluation en éducation » (Mottier Lopez et Figari, 2012a) et vient déployer une réflexion sur trois crises du sens. Une première, à proprement parler théorique, générée par un « éclatement des explications possibles » (2012, p. 10) de l'évaluation. Une deuxième que nous qualifions d'épistémique, porte sur la réduction des activités évaluatives à des activités de jugement non scientifiques : ce faisant, toute tentative d'investiguer et de modéliser scientifiquement l'évaluation serait annulée d'office, celle-ci ne pouvant prétendre à ce statut. Une dernière que nous qualifions d'éthique porte sur la non-neutralité de l'évaluation « en tant qu'activité finalisée, située dans un contexte social, historique et politique donné » (2012, p. 11).

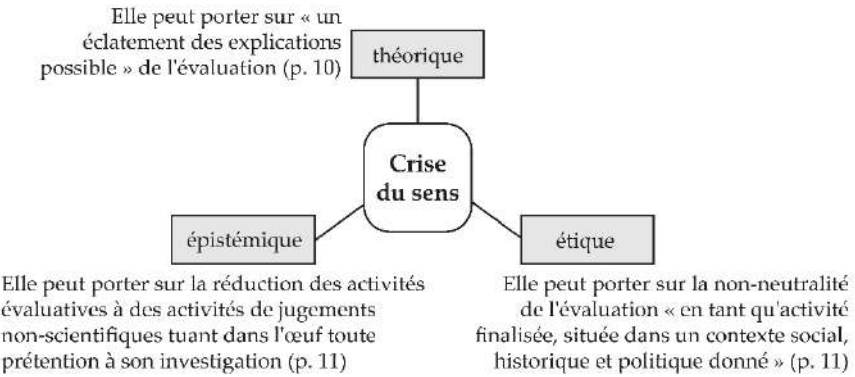


Figure 1 : Trois crises du sens en évaluation d'après Mottier Lopez et Figari (2012)

Avec ces deux contributions, nous identifions deux sortes de crises. Une première porterait sur ce qui qualifie la valeur dans l'évaluation. Si ce dernier concept ou cette activité n'est pas stabilisé, les fondations sociales de l'évaluation s'en voient instables à leur tour. La crise, que nous osons qualifier d'anthropologique, vient reposer la question de ce qu'est l'évaluation de façon radicale : que nomme l'activité qui constitue et repose sur la valeur ? Une seconde porterait sur ce qui qualifie le sens de l'évaluation : une crise épistémologique. Au moins trois sens-directions étaient identifiés alors. Aujourd'hui, ils sont élargis à leur tour par l'Intelligence Artificielle, notamment générative, et les *Learning Analytics* venant mettre en crise le rapport de l'humain-évaluateur ou évaluatrice aux informations à prélever ou à juger dans l'acte même d'évaluer (Detroz, 2024), voire à ses pouvoirs comme ses limites, interrogeant par-là même sa responsabilité (Roblez, 2025). Ces enjeux contemporains se posent comme une quatrième crise du sens : celui de l'humanité de nos activités et des tentations techniques ou technologiques à déléguer sans reléguer notre responsabilité. Cette quatrième orientation nous paraît reboucler avec la crise anthropologique du début, puisqu'au fond : qui évalue quoi ? Nous posons ces deux types de crises comme ce qui constitue l'axe horizontal de l'évaluation : sa dimension anthropologique et sa quête de sens constituent sa base.

Mais après tout, la tentation technologique n'est-elle pas constitutive de l'évaluation en tant que telle ? En 2014, Demailly avertissait qu'évaluer devait rencontrer une « salutaire résistance » (2014) à son irrésistibilité. Car, dans le même temps que les questionnements et les recherches se

déroulaient - et continuent à se dérouler - deux autres crises apparaissent, sur l'axe vertical de l'évaluation, qui représente sa structure.

La première est sociétale et vient montrer les limites du néolibéralisme et du capitalisme, de leurs valeurs dans l'éducation (Fabre et Gohier, 2015). L'évaluation fricote bien avec les fondements néolibéraux : elle pourrait même en être un archétype (Roblez et Progin, soumis). Charles Hadji identifie une structure complète d'une évaluation au service du marché : la « néo-évaluation » (Hadji, 2021). Pour veiller à sa bonne hégémonie, la Nouvelle gestion publique doit veiller à ce que les choses soient harmonisées, simplifiées, déposées dans un même cadre analytique, afin de mieux les gérer (Amable, 2023). Or les limites sont atteintes depuis bien longtemps : par le rappel du vivant et de la matière, par les violences institutionnelles et d'injustices tant sociales qu'environnementales comme par les reproductions des dominations, tout ceci est aujourd'hui dénoncé, combattu, repositionnant les structures de nos sociétés. Cette crise sociétale interroge à son tour les limites de l'évaluation et son sens, relatif à une certaine vision de l'humain, de ce qui compte, de la civilisation. Elle est symbolisée en haut de l'axe vertical puisqu'elle retombe pour ainsi dire sur la base de l'axe horizontal.

La seconde crise est dans notre rapport aux savoirs. Elle est liée à la précédente, dans la mesure où la société vit et se transforme, se pluralise. Les frontières entre le savoir sachant et les savoirs se faisant, les savoirs institués et les savoirs destitués, les savoirs adossés et ceux crus se complexifient, s'amalgament dans les discours ou les positions des personnes. Aux « guerres des paradigmes » de Gage (1989 citées dans Younès, 2020) se succèdent des conflits de valeurs interrogeant les fondements de ces structures scientifiques : est-il toujours possible, voire bon, d'évaluer les savoirs des élèves de différents pays à des fins de classement, sans prendre en considération les crises des métiers, des services, des foyers ou des personnes elles-mêmes ? Est-il éthiquement juste d'utiliser une évaluation à des fins de régulation sans concerter l'autre, sans savoir s'il ou si elle sait, cette autre personne, elle aussi ce dont il s'agit, les fins poursuivies, le cheminement opérationnel à conduire ? Si nous sommes de plus en plus attentifs et attentives aux émotions, aux affects, aux histoires de vie dans les modélisations puis usages des évaluations, les apports des théories alternatives n'en sont encore qu'à leur balbutiement : que ce soit les *Genders studies* (ne se limitant pas qu'au biais de genre), les études féministes, les *Subalterns studies* ou *Climates studies*… ces différentes épistémologies critiques contemporaines interrogent toutes nos rapports aux savoirs et les degrés d'implication et d'historicisation des connaissances, rappelant qu'elles sont concomitantes d'une époque, de ses

mœurs et de ses agents. Cette crise de notre rapport aux savoirs se pose en bas de l'axe vertical, car elle prend du temps, elle consolide ou fragilise les fondations de l'axe horizontal, de la base de l'évaluation qui recherche sa structure scientifique, sans perdre ses fondements éthiques.

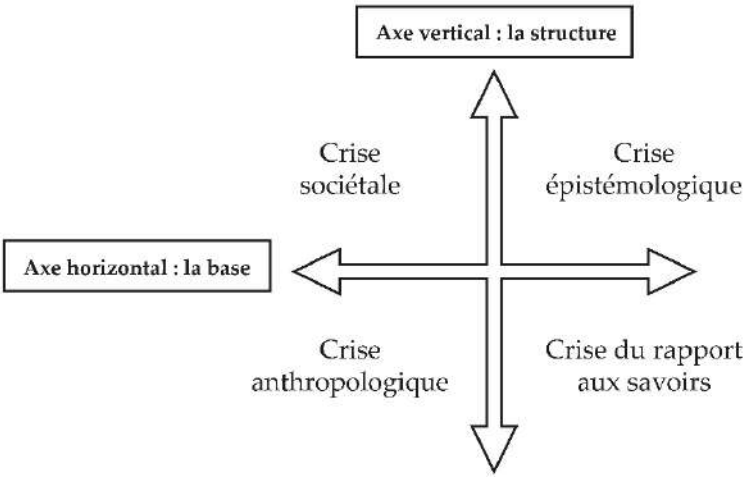


Figure 2 : Quatre grandes crises dans l'évaluation

IV. « Où vont les valeurs ? » Un moment critique dans notre société de l'évaluation

La question de Jérôme Bindé et des chercheur·es, philosophes, politiques pour l'Organisation des Nations unies pour l'éducation, la science et la culture, nous semble pouvoir se reposer pour ce qui concerne l'évaluation et ses transformations : *Où vont les valeurs ?* (Bindé, 2004). Car ce que l'évaluation montre, ce n'est pas tant un manque de valeurs dans notre monde que probablement plus un « élargissement » de leur nombre et de leurs qualifications, de leurs nuances ou de leurs formes de manifestations : ce que Weber appelait un « polythéisme des valeurs » (1963). « Il n'y a pas tant crise des valeurs - nous n'en manquons pas - que crise du sens même des valeurs, et de l'aptitude à se gouverner. La question urgente est donc de savoir comment s'orienter parmi les valeurs » (Mastuura, 2004, p. 10).

Notre hypothèse est que nous nous trouvons à un carrefour de polycrises nous mettant face aux liaisons et aux liens qu'ont l'ensemble de nos activités, dont l'évaluation. Dans un monde hyperconnecté, les chocs et perturbations se retrouvent exacerbées, par « résonance » (Rosa, 2018) : les problématiques ne se pensent plus seules ou coupées des autres difficultés ou problèmes rencontrés. Ce faisant, évaluer est un analogon : une forme de détecteur des liens résonants entre eux par la mise à jour de valeurs - celles portées par l'institution, celles de la personne qui évalue, celles de la personne dite évaluée, celles des autres (imaginaires et réels). Et qui dit mise à jour, dit risque d'ombrage, de mise à l'ombre, de zones ou de valeurs. C'est la raison pour laquelle, plus précisément, notre hypothèse est que l'évaluation fait partie de ces activités catalysatrices de ces polycrises. En mettant à jour les valeurs d'une production, d'une performance, d'une situation…, nous engageons réciproquement leur consistance, leur propre valeur en somme, en l'exerçant. Lecointe élargissait à son tour la relation connue dans les théories de la référentialisation (Figari et al., 2014 ; Mottier Lopez, 2017) en rajoutant un volet au référent et aux référés d'évaluation : celui de la « référence » (Lecointe, 1997).

> Ce que, dans la triple articulation de l'évaluation, j'ai appelé la dimension « référence » par opposition et complément aux dimensions « référé » et « référent » représente assez bien la trace permanente, le constant mouvement « d'axiologisation », ou, en son sens le plus fort, la constante évaluation, le permanent mouvement de décision de valeur (Lecointe, 1997, p. 162).

La « référence », reprise plus tard par Vial pour évoquer le « système de références » (Vial, 2009) pour fonder sa problématique, constitue le support idéologique et/ou philosophique dans lequel s'ancre le référent et les référés, leur donne une « justification » (Boltanski et Thévenot, 2008) sur le versant social, et leur donne une contenance éthique sur le versant réflexif. Pour Lecointe, une évaluation fabrique des valeurs en utilisant des structures - normatives, idéologiques, fonctionnelles, etc. - existantes et en élaborant d'autres se faisant. L'expression contemporaine prisée par la jeunesse – « T'as la réf ? » - donne à voir la constitution de toute évaluation : elle est d'abord et avant tout sociale, connectrice de références entre les individus. « Comment s'orienter parmi les valeurs », pour reprendre Mastuura, qui se confrontent, se multiplient et se diversifient d'un côté, tandis que d'un autre, la société se cherche, se critique, se transforme dans un univers de comparaisons sociales, d'inégalités accentuées, de santés délétères voire de souffrances ? Au fond, l'évaluation est partout dans notre société : se juger, se comparer, critiquer sont des

actions communes. Évaluer, c'est s'orienter : peut-on imaginer des évaluations autres que « folie » (Abelhauser et al., 2011), pernicieuses (Hadji, 2012) ou « tyranniques » (Del Rey, 2013), sources de « menaces » (Butera et al., 2011) ? Il nous semble que oui.

V. Et dans 40 ans ? Bref essai de prospective axiologique pour l'évaluation en éducation et en formation

L'avantage, lorsqu'on étudie les crises et qu'on utilise ce concept pour entrer dans l'analyse des phénomènes, c'est qu'il se constitue un ascétisme mobilisateur pour son après-coup. Car tout phénomène crisique révèle quelque chose : c'est que ce qui est en crise est en vie. Une vie biologique, matérielle ou figurée : mais en vie quand même. Par conséquent, la crise ne marque pas un arrêt, un *statu quo* : elle est « ce dont il faut partir ou repartir, ce dont le statut doit être réélaboré » (Revault d'Allonnes, 2012, p. 110). Tentons donc un bref exercice de prospective des valeurs dans notre paysage scientifique.

Il nous semble concevable que les réflexions et interrogations soulevées lors du dixième anniversaire de l'ADMEE-Europe coïncident encore aujourd'hui. L'ordre économique certes a eu le vent en poupe ; mais pour autant, l'ordre éthique ne s'est pas fait totalement occulter par celui-ci. Au contraire, l'ordre économique s'est approprié ce dernier, par l'utilisation de valeurs éthiques ou morales, à des fins d'évaluation institutionnalisée : par exemple, les comités d'évaluation éthique de la recherche (Allaire, 2022) se présentent comme un espace concerté pour décider du respect des personnes ou des informations, lors d'une recherche. Un autre exemple concerne les démarches de prévention des risques au travail (Caroly et al., 2015), utilisées aux seules fins de régulation de situations délicates, c'est-à-dire dans l'après-coup… de crises au travail. Cela signifie donc que tout un empan existe en dehors de l'ordre économique ; voire que l'ordre éthique peut venir à son tour interroger et orienter l'ordre économique. Car après tout, « l'économie est une science morale » (Sen, 2003). Des recherches vont proposer des jonctions entre les deux « types de questionnements » (Figari, 2001) posés par Lecointe et Triby, pour ainsi assumer la nature élaboratrice de valeurs de toute évaluation. L'une des façons sera d'observer les agents animateurs des crises, puisque « Ce qu'on appelle 'crise' est toujours une construction sociale. Qu'elle repose ou non sur des faits, elle a besoin d'agents qui la légitiment » (Fassin, 2023, p. 32).

Gérard Figari identifiait ainsi un « village des évaluateurs » Figari, 2012, p. 223 agencé par

> 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 savoir 'sociaux', 'stratégiques', 'd'action', 'de jugement', 'profanes' que l'évaluation contribue à construire (2012, p. 223).

Ce « village » ou ces « villages » des deux côtés de l'Atlantique nous semblent jouer le jeu de cette « culture scientifique » invoquée. Réflexivement, cette culture appelle à identifier les gestes, non seulement professionnels, mais aussi plus généralement symboliques, psychologiques, affectifs participant à attribuer de la valeur. Les quatre crises épistémologiques nous ramènent à notre humanité : non seulement elles appellent un regard critique sur nos responsabilités, mais elles invitent à ne pas s'y limiter, puisque si on suit Martuccelli, la responsabilité peut servir de pendant à l'assèchement des liens sociaux en mettant l'individu comme celui qui assume seul (2014). S'inspirer du vivant, dans une sociologie élargie à son tour (Lahire, 2025) incluant « l'autre pas qu'humain », pour comprendre les constructions, maintiens et défections de nos valeurs, pour ré-anthropologiser d'une part ce qui peut avoir tendance à ne passer que pour une technique ; et se reconnecter d'autre part à ce qui vit, et co-construire nos vies en lucidité.

Les recherches pourraient porter sur la « robustesse » (Hamant, 2023) des processus et des activités d'évaluation : ceux et celles qui tiennent, non par contrainte ou arbitraire de règles, mais par leur homéostasie avec « le reste » pour reprendre Vial (2009), ce qui ne rentre pas dans les cases mais qui se trouve « là ». La « logique de contrôle » (Vial, 2009) étant connue et identifiée, ne serait pas écartée ou abandonnée par déni, mais réfléchie pour être encore davantage partagée, répartie, tournante ou erratique : un « agir évaluationnel » (Hadji, 2021 ; Lecointe, 1997) favoriserait les échanges et les transactions sur les effets sociaux, l'évaluation intégrant par conséquent toute une gamme de références « dialogiques » (Figari et Gremion, 2020). Comment discuter de nos valeurs différentes, voire contradictoires ? À quelles fins ? La dimension politique de l'évaluation côtoierait la dimension technique, qui s'affinerait à partir de ressources différenciées, mais non écocidaires.

Enfin, les recherches pourraient modéliser davantage une évaluation crisologique. Au carrefour de différentes disciplines, elle générerait des activités de fictions pour mieux comprendre et agir non plus sur la réalité, mais dans la réalité. Les fictions pourraient être utilisées comme des

« métaréférents », en positionnant des références différenciées dans leur temporalité : la conscience des références perceptibles, actuelles et d'hier, n'enclave pas les espoirs et les idéaux d'alternatives. Les fictions-métaréférents positionneraient le désirable dans quelque chose qui peut se désirer aujourd'hui, impliquant de réfléchir aux moyens par les fins à discuter ; et à critiquer les fins par les moyens à solliciter. Ce faisant, des références partagées, démocratiques, pourraient se co-constituer : non pas par un seul pouvoir, qu'il soit scientifique, politique ou une autre forme hégémonique. Mais par nos imaginaires et la difficile - mais réjouissante - capacité à agir dans un monde incertain, avec nos contradictions.

VI. Bibliographie

Spéciale « crise » dans nos publications

Demailly, L. (2006). En Europe : L'évaluation contre la crise des systèmes scolaires, l'évaluation en crise. *Éducation et sociétés*, *17*(1), 105-120. Cairn.info. https://doi.org/10.3917/es.017.0105

Demailly, L. (2013). Crise, réagencements identitaires et déprofessionnalisation dans les métiers de l'éducation : Entretien avec L. Demailly réalisé par F. Salane. *Recherche & formation*, *74*, 115-124. https://doi.org/10.4000/rechercheformation.2152

Detroz, P., Tessaro, W. et Younès, N. (2020). *Evaluer en temps de pandémie* (e-JIREF). ADMEE-Europe. https://journal.admee.org/index.php/ejiref/issue/view/16

Deux articles :

Rached, P., Gharib, Y. et Constantin, S. (2020). Entre le cognitif et l'émotionnel... Quelle évaluation à distance en situation de crise ? e-JIREF, 1(Hors-série), 223-230.

Roblez, A., Biancarelli, L. et Gremion, C. (2020). Dialogue transdisciplinaire sur l'évaluation en temps de crise—Critiques et Perspectives. *e-JIREF*, *1*, 237-243.

Mottier Lopez, L. et Figari, G. (2012). *Modélisations de l'évaluation en éducation : Questionnements épistémologiques*. De Boeck. https://www.unige.ch/fapse/editions/files/8215/6094/6047/RE16.pdf

Bibliographie générale

Abbott, A. (2009). 11. À propos du concept de Turning Point. Dans *Bifurcations* (p. 187-211). La Découverte ; Cairn.info. https://doi.org/10.3917/dec.bessi.2009.01.0187

Abelhauser, A., Gori, R. et Sauret, M. J. (2011). *La folie évaluation : Les nouvelles fabriques de la servitude*. Avec L'Appel des appels. Essai Mille et une nuits.

Allaire, S. (2022). Quelques idées pour aplanir les tensions entourant le travail des comités d'éthique. *Revue des sciences de l'éducation*, *48*(3), 1103277ar. https://doi.org/10.7202/1103277ar

Amable, B. (2023). *Le néolibéralisme*. Que sais-je ?

Ardoino, J. (2001). Conclusion. De la méthodologie à l'épistémologie. Dans G. Figari et M. Achouche (dir.), *L'activité évaluative réinterrogée* ; (p. 227-231). De Boeck Supérieur. DOI: 10.3917/dbu.figar.2001.01. https://shs.cairn.info/l-activite-evaluative-reinterrogee--9782804136147-page-227?lang=fr

Arendt, H. (2013). *La crise de la culture : Huit exercices de pensée politique* (traduit par P. Lévy). Gallimard.

Bernard de Raymond, A. (2023). La société comme « système complexe » : Vers une physique du monde social ? *Revue d'anthropologie des connaissances*, *17*(3). https://doi.org/10.4000/rac.30519

Bindé, J. (2004). *Où vont les valeurs ? Entretiens du XXIe siècle II* (Unesco). UNESCO. https://archive.org/details/isbn_9782226142450

Boltanski, L. et Thévenot, L. (2008). *De la justification : Les économies de la grandeur.* Gallimard.

Butera, F., Buchs, C. et Darnon, C. (2011). *L'évaluation, une menace ?* Presses universitaires de France.

Caroly, S., Simonet, P. et Vézina, N. (2015). Marge de manœuvre et pouvoir d'agir dans la prévention des TMS et des RPS : *Le travail humain*, *78*(1), 1-8. https://doi.org/10.3917/th.781.0001

de Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, *XV*, 25-37.

Del Rey, A. (2013). *La tyrannie de l'évaluation*. La Découverte.

Demailly, L. (2006). En Europe : L'évaluation contre la crise des systèmes scolaires, l'évaluation en crise. *Éducation et sociétés*, *17*(1), 105-120. Cairn.info. https://doi.org/10.3917/es.017.0105

Demailly, L. (2014). Préface. La salutaire résistance à l'irrésistible évaluation. Dans J. Morrissette et M.-F. Legendre, *Enseigner et évaluer : Regards sur les enjeux éthiques et sociopolitiques* (p. XIII-XVIII). Hermann PUL, Presses de l'Université de Laval.

Detroz, P. (2024). *Conférence d'ouverture : Plongée en mer de données, à la recherche de trésors enfouis* [Conférence]. Conférence d'ouverture du REF 2024, Fribourg. https://orbi.uliege.be/handle/2268/320349

Diamond, J. M. (2009). *Effondrement : Comment les sociétés décident de leur disparition ou de leur survie* (traduit par A. Botz et J.-L. Fidel). Gallimard.

Fabre, M et Gohier, C. (2015). *Les valeurs éducatives au risque du néo-libéralisme*. Presses universitaires de Rouen et du Havre.

Fassin, D. (2023). *Sciences sociales par temps de crise*. Collège de France éditions.

Figari, G. (2001). Conclusion. L'activité évaluative et la recherche : Rapports en débat. Dans M. Achouche (dir.), *L'activité évaluative réinterrogée* (p. 361-364). De Boeck Supérieur. http://www.cairn.info/l-activite-evaluative-reinterrogee--9782804136147.htm DOI: 10.3917/dbu.figar.2001.01

Figari, G. (2012). Configurations pour une épistémologie de l'évaluation. Dans L. Mottier Lopez et G. Figari (dir.), *Modélisations de l'évaluation en éducation. Questionnements épistémologiques* (p. 211-224). De Boeck.

Figari, G. et Achouche, M. (1997). Dix années de travaux de recherche en évaluation (1986-1996) : Bilan tiré des travaux du X colloque ADMEE-Europe (septembre 1996). *Mesure et évaluation en éducation*, *19*(3), 5. https://doi.org/10.7202/1091394ar

Figari, G. et Achouche, M. (2001). *L'activité évaluative réinterrogée*. De Boeck Supérieur. http://www.cairn.info/l-activite-evaluative-reinterrogee--9782804136147.htm (DOI: 10.3917/dbu.figar.2001.01)

Figari, G. et Gremion, C. (2020). Conclusion : Vers une évaluation dialogique ? Dans N. Younès, C. Gremion et E. Sylvestre (dir.), *Evaluations, sources de synergies* (p. 250-267). Presses de l'ADMEE.

Figari, G., Remaud, D. et Tourmen, C. (2014). *Méthodologie d'évaluation en éducation et formation ou L'enquête évaluative*. De Boeck.

Foucault, M. (2008). *L'archéologie du savoir* (Œuvre originale publiée en 1969). Gallimard.

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

Hadji, C. (2021). *Le défi d'une évaluation à visage humain. Dépasser les limites de la société de performance*. ESF Editeur.

Hamant, O. (2023). *Antidote au culte de la performance : La robustesse du vivant*. Gallimard.

Heurtaux, J., Renault, R. et Tarragoni, F. (2023). Etats de crise. Dans *Etats de crise* (p. 9-27). ENS Éditions. https://journals.openedition.org/traces/15059

Husserl, E. (1976). *La crise des sciences européennes et la phénoménologie transcendantale* (traduit par G. Granel). Gallimard.

Koselleck, R. (2006). Crisis. *Journal of the History of Ideas*, *67*(2), 357-400. https://doi.org/10.1353/jhi.2006.0013

Kuhn, T. S. (2018). *La structure des révolutions scientifiques*. Flammarion.

Lahire, B. (avec Sanseigne, F. et Flandrin, L.). (2025). *Vers une science sociale du vivant*. La Découverte.

Laveault, D. (2014). Mesure sans démesure : La contribution de Jean Cardinet aux méthodologies de la mesure et de l'évaluation en éducation. *Mesure et évaluation en éducation*, *31*(2), 5-17. https://doi.org/10.7202/1025004ar

Lecointe, M. (1997). *Les enjeux de l'évaluation*. L'Harmattan.

Martuccelli, D. (2014). La responsabilisation, nouvelle forme de domination. Dans *L'Individu contemporain* (p. 206-208). Éditions Sciences Humaines. Cairn.info. https://doi.org/10.3917/sh.molen.2014.01.0206

Mastuura, K. (2004). Préface. Dans J. Bindé (dir.), *Où vont les valeurs ? Entretiens du XXIe siècle II* (Unesco, p. 9-11). UNESCO. https://archive.org/details/isbn_9782226142450

Merle, P. (2018). *Les pratiques d'évaluation scolaire : Historique, difficultés, perspectives*. PUF.

Morin, E. (2012). Pour une crisologie : *Communications*, *91*(2), 135-152. https://doi.org/10.3917/commu.091.0135

Mottier Lopez, L. (2017). Une modélisation pour appréhender la référentialisation dans l'évaluation des apprentissages des élèves. Dans P. Detroz, M. Crahay et A. Fagnant (dir.), *L'évaluation à la lumière des contextes et des disciplines* (p. 169-192). De Boeck Supérieur.

Mottier Lopez, L., Blanc, Dechamboux, L. et Tobola Couchepin, C. (2017). Les héritages de Jean Cardinet : Regards à partir de trois recherches doctorales sur l'évaluation des apprentissages des élèves en classe. *e-JIREF*, *3*(3). https://doi.org/10.48782/vgpr1377

Mottier Lopez, L. et Figari, G. (2012a). Modèles et modélisations face à une crise du sens de l'évaluation en éducation. Dans L. Mottier Lopez et G. Figari (dir.), *Modélisations de l'évaluation en éducation : Questionnements épistémologiques* (1e éd., p. 7-23). De Boeck Supérieur.

Mottier Lopez, L. et Figari, G. (2012b). *Modélisations de l'évaluation en éducation : Questionnements épistémologiques*. De Boeck. https://www.unige.ch/fapse/editions/files/8215/6094/6047/RE16.pdf

Pelletier, L. (1971). La notion d'évaluation. *Education permanente*, *9*, 7-19.

Popper, K. R. (1998). *La connaissance objective* (traduit par J.-J. Rosat.). Flammarion.

Revault d'Allonnes, M. (2012). *La crise sans fin essai sur l'expérience moderne du temps*. Éditions du Seuil.

Ricœur, P. (2019). 9. La crise : Un phénomène spécifiquement moderne ? Dans *Politique, économie et société* (p. 165-196). Le Seuil ; Cairn.info. https://www.cairn.info/politique-economie-et-societe-ecrits-et-conference--9782021419429-p-165.htm

Roblez, A. (2024a). Les pratiques évaluatives et le HCERES : réflexions pour dessiner d'autres frontières. Dans B. Mocquet, C. Pélissier et P.-M. Riccio (dir.), *Pratiques d'évaluation et transformations.*

Management des Technologies Organisationnelles (TRANSVALOR, p. 181-196). Presses des Mines.

Roblez, A. (2024b, février 7). *C'est la crise : Un contexte de valeurs en tension dans l'évaluation* [Conférence]. Cycle de conférences 2023-24 de l'ADMEE-Suisse, Lyon. https://hal.science/hal-04531800/

Roblez, A. (2024c). L'évaluation comme une activité illustratrice de crises : Une place pour l'inanticipable ? *Questions vives recherches en éducation, N° 41*. https://doi.org/10.4000/14hr4

Roblez, A. (2025, février 25). *Intelligence évaluative et évaluation artificielle : Réflexions sur la relation entre l'IA et l'évaluation* [Webinaire du Réseau de la pédagogie (ICL)]. https://univ-catholille.hal.science/hal-04967764

Roblez, A. et Progin, L. (soumis). Les cadres scolaires à l'épreuve de la néo-évaluation : Tensions, ajustements et bricolage pour s'émanciper ? *REE*.

Rosa, H. (2018). *Résonance : Une sociologie de la relation au monde* (traduit par S. Zilberfarb). La Découverte.

Sartre, J.-P. (1970). *L'existentialisme est un humanisme*. Nagel.

Sen, A. (2003). *L'économie est une science morale*. La Découverte.

Servigne, P. et Stevens, R. (avec Cochet, Y.). (2015). *Comment tout peut s'effondrer : Petit manuel de collapsologie à l'usage des générations présentes*. Éditions du Seuil.

Tenret, É. (2021). Chapitre 9. L'orientation à l'heure des algorithmes. Sélection automatisée, vécu de l'orientation et sentiment de justice parmi les étudiants. Dans V. Cohen-Scali (dir.), *Psychologie de l'orientation tout au long de la vie* (p. 171-192). Dunod. Cairn.info. https://doi.org/10.3917/dunod.cohen.2021.01.0171

Vial, M. (2009). *Se former pour évaluer—Se donner une problématique et élaborer des concepts* (2e éd.). De Boeck Supérieur.

Weber, M. (1963). *Le savant et le politique*. Union Générale d'Éditions. http://classiques.uqac.ca/classiques/Weber/savant_politique/Le_savant.html

Younès, N. (2020). *Vers une évaluation écologique dans l'enseignement supérieur : Dispositifs et transformations en jeu* [HDR, université de

Lorraine]. https://hal.archives-ouvertes.fr/tel-03032417/document

Table des matières

Introduction **13**

Chapitre 1. 40 ans , cela se fête ! **21**
Alain Bouvier

Chapitre 2. MEE et e-JIREF : Parcours et enjeux de deux revues scientifiques francophones **29**
Carla Barroso da Costa, Cathy Perret, Nathalie Younès

I. Introduction 29
II. L'essor de MEE et d'e-JIREF dans la publication scientifique francophone en évaluation 30
II.a. Mesure et Évaluation en Éducation (1978) - « Le navire est lancé » 30
II.b. e-JIREF (2015) – Naissance d'une revue aux multiples défis 32
III. L'accès ouvert pour la qualité et le partage démocratique des résultats de la recherche 33
IV. Mutations et adaptations au fil du temps : des enjeux en commun 34
IV.a. Les enjeux de visibilité à l'ère du numérique 35
IV.b. Les enjeux de la publication scientifique en français à l'heure de l'anglicisation 36
IV.c. Une activité éditoriale exigeante reposant sur l'engagement des membres de la communauté 37
V. Et quel est l'avenir des publications scientifiques ? 38
VI. En guise de conclusion 40
VII. Bibliographie 42

Chapitre 3. L'éternel débat des notes à l'école **45**
Charles Hadji

I. Introduction 45
II. Culture de la performance scolaire et bienveillance éducative 46
III. Evaluer pour accompagner ou pour classer ? 47
IV. S'entendre sur les objectifs du système scolaire 48
V. Bibliographie indicative 49

Chapitre 4. Y a-t-il encore un avenir pour l'évaluation ? **51**
Jean-Marie De Ketele

I. Introduction 51
II. L'évaluation sous le feu des critiques 52

III. L'avenir est-il dans l'évaluation qualitative ? 55
IV. Un besoin d'élargir le champ de l'évaluation 58
V. Ainsi donc, l'évaluation a plus que jamais un avenir 62
VI. Bibliographie 65

Chapitre 5. L'évaluation formative et ses reconfigurations conceptuelles 71
Lucie Mottier Lopez
I. L'évaluation formative, une impulsion à la création de l'ADMEE-Europe 71
II. Apports des travaux francophones jusque dans les années 2000 : l'évaluation formative « élargie » 72
III. Conceptualisations renouvelées de l'évaluation formative dans les travaux anglophones dès les années 2000 75
IV. Une évaluation continue pour apprendre durablement 78
V. Perspectives critiques 80
VI. Bibliographie 83

Chapitre 6. Evolution, tensions et perspectives de l'évaluation pour toutes et tous 89
Françoise Pasche Gossin
I. Introduction 89
II. Une lente émergence de l'évaluation 91
II.a. D'une pratique d'évaluation informelle à une évaluation uniforme et standardisée au XIXe siècle 92
II.b. Développement des tests et des mesures à une remise en question et diversification de l'évaluation au XXe siècle 93
II.c. D'une évaluation-soutien d'apprentissage à une évaluation pour toutes et tous au XXIe siècle 94
III. L'évaluation pour toutes et tous au cœur de tensions dialectiques 96
III.a. Entre uniformisation et différenciation 96
III.b. Entre normalisation et émancipation 97
III.c. Entre bienveillance et exigence 98
III.d. Entre accompagnement et certification 99
IV. Perspectives de l'évaluation pour toutes et tous 100
V. Conclusion 102
VI. Bibliographie 103

Chapitre 7. RCPE : des recherches collaboratives sur l'évaluation en éducation, d'une approche initialement marginale à une nouvelle alliance entre recherche et formation 107
Lucie Mottier Lopez, Céline Girardet
I. Création du réseau RCPE : recherches collaboratives sur les pratiques évaluatives 107
II. Analyse des articles faisant partie des numéros spéciaux RCPE, et de ceux qui ont été publiés dans les revues de l'ADMEE 110
II.a. Constitution de la base de données et description 111
II.b. Analyse de contenu des 35 articles 112
III. Résultats de l'analyse des articles retenus 112

III.a. Articles dont l'objet d'étude est la RC 113
III.b. Articles qui étudient le rapport entre RC et évaluation 114
III.c. Articles dont la RC est la démarche méthodologique adoptée 115
III.d. Perspectives 119
IV. Bibliographie 120
V. Annexe 1 122

Chapitre 8. De l'évaluation comme objet d'étude à l'évaluation comme pratique de recherche 125
L'essor de la recherche-intervention évaluative, défis et perspectives
Lucie Aussel, Isabelle Bertolino, Dominique Broussal, Christelle Chauffriasse, Manon Delbreil, Pauline Hédacq 125
I. Introduction 125
II. Évaluation, quand tu nous tiens 126
III. Caractérisation de la recherche-intervention évaluative au sein des démarches de recherche-intervention 129
IV. L'appréhension du pouvoir au sein de la RIE 133
IV.a. Considérations liminaires sur la question du pouvoir 133
IV.b. La RIE en tant que dispositif : espace de liberté et de contraintes 134
V. Des recherches doctorales à l'épreuve de la RIE : quatre docteures témoignent 138
V.a. Témoignage d'Isabelle Bertolino 139
V.b. Témoignage de Christelle Chauffriasse 141
V.c. Témoignage de Manon Delbreil 142
V.d. Témoignage de Pauline Hédacq 144
VI. Conclusion 146
VII. Bibliographie 146

Chapitre 9. Biais du jugement évaluatif en contexte scolaire 153
Dylan Dachet, Ariane Baye, Dominique Lafontaine et Marc Demeuse
I. Introduction 153
II. Émergence de la docimologie descriptive 154
III. Développement d'une docimologie critique 155
IV. Docimologie du jugement au sein des classes 157
V. Études et méta-analyse sur l'effet de halo 160
VI. Apports de la psychologie 162
VII. Peut-on lutter contre les biais évaluatifs ? 164
VIII. Conclusion 165
IX. Bibliographie 167

Chapitre 10. La mesure dans la langue de Molière 177
Etat des lieux et avenues à envisager
Dominique Casanova et Sébastien Béland 177
I. Introduction 177
II. Qu'est-ce que la mesure ? 178
III. Le contenu de la revue Mesure et évaluation en éducation : le cas de la mesure 180

III.a. Les thèmes étudiés dans la revue MEE 181
IV. Quoi de neuf en mesure ? 187
IV.a. Une explosion des travaux mobilisant l'intelligence artificielle (IA) 188
IV.b. Les modèles de réseaux pour mieux comprendre les relations de dépendance entre les items de vos épreuves 189
IV.c. La validité : vers de nouveaux standards en 2027 ! 192
IV.d. Les modélisations bayésiennes se démocratisent (enfin) 193
IV.e. Les modèles de mesure modernes sont toujours en développement 193
V. Conclusion 194
VI. Bibliographie 194

Chapitre 11. Compétences non académiques : malgré les difficultés de conceptualisation et de mesure, un intérêt majeur pour les chercheurs et les praticiens en éducation 201
Sophie Morlaix
I. Introduction 201
II. Les compétences non académiques : des dénominations et des définitions multiples dans des recherches de plus en plus nombreuses et qui cherchent à les mesurer 202
III. Une complémentarité pourtant essentielle avec le développement des compétences cognitives 207
IV. Au-delà des compétences non académiques développées par les élèves, l'importance de celles développées par les enseignants pour contribuer au processus d'apprentissage 209
V. Compétences non académiques, marché du travail et bénéfices sociétaux 211
VI. Pour conclure… 214
VII. Quelques éléments de bibliographie 214

Chapitre 12. La menace de l'évaluation : 40 ans de recherche 217
Fabrizio Butera
I. Introduction 217
II. Les précurseurs 218
III. Les années 1980 219
IV. Les années 1990 220
V. Les années 2000 221
VI. Les années 2010 223
VII. Les années 2020 et le futur 224
VIII. Bibliographie 225

Chapitre 13. Comment « évaluation » et « didactique » s'articulent-elles dans deux champs de recherche anglophones majeurs : les Pedagogical Content Knowlegde et l'Assessment Literacy ? 231
Évaluation et didactique dans le champ des PCK et de l'AL
Céline Lepareur, Annick Fagnant et Raphaël Pasquini
I. Introduction 231
II. La place variable de l'évaluation dans les modèles PCK pour l'enseignement des sciences et des mathématiques 233
II.a. Des composantes de PCK contrastées dans les modèles pour l'enseignement des sciences 234
II.b. La place discrète de l'évaluation dans les modèles de PCK en mathématiques 238
III. L'inscription des PCK dans l'Assessment Literacy 242
III.a. Une modélisation de l'AL où les PCK jouent un rôle central 244
III.b. Un exemple de recherche croisant PCK et AL 246
IV. Discussion et perspectives 247
V. Bibliographie 250
VI. Annexes 255

Chapitre 14. Évaluer la professionnalisation 257
Apport de 40 ans de publication de l'ADMEE
Christophe Gremion
I. Introduction 257
II. Quelques éléments conceptuels 258
II.a. De quelle professionnalisation parle-t-on ? 258
II.b. Professionnalisation de qui ou de quoi ? 259
II.c. Professionnalisation pour quoi ? 260
II.d. Quels indicateurs de professionnalisation ? 261
III. Méthodologie 262

IV. Résultats 263
IV.a. Les objets concernés par la professionnalisation 263
IV.b. Les définitions de la professionnalisation 265
IV.c. Les indicateurs mobilisés pour observer la professionnalisation 266
V. Discussion et conclusions 270
VI. Bibliographie 272

Chapitre 15. Réseau thématique international Reconnaissance, valorisation et validation des acquis de l´expérience de l´ADMEE-Europe 277
Dynamique d'un groupe de recherche et orientations épistémologiques, théoriques et empiriques des investigations
Carmen Cavaco et Pascal Lafont
I. Introduction 277
II. Construction et dynamique d'un groupe thématique de recherche 279
II.a. Constitution évolutive du groupe de recherche 279
II.b. Une activité scientifique régulière et structurante 280
III. Productions scientifiques et éléments d'analyse 282
III.a. Productions scientifiques 282
III.b. Questionnements épistémologiques et orientations des recherches 285
III.c. Dimensions éthiques 287
IV. Conclusion 288
V. Bibliographie 289

Chapitre 16. Évaluer en situation de travail, un renouvellement des questions d'évaluation des compétences 293
Patrick Mayen
I. Introduction 293
II. Évaluer en situation de travail, un renouvellement des questions d'évaluation des compétences 294
III. Les conditions « institutionnelles » de l'évaluation 296
IV. Analyse du travail et analyse de la formation : un processus de référentialisation 297
V. Évaluer le potentiel formatif des situations 299
VI. Evaluer les compétences 300
VII. Conclusion 302
VIII. Bibliographie 303

Chapitre 17. Quand dire l'évaluation, c'est faire apprendre 305
La performativité des discours sur l'évaluation
Marc Romainville
I. Introduction 305
II. Missions, fonctions et effets de l'évaluation 307
III. L'évaluation comme acte perlocutoire 308
IV. Des effets perlocutoires variés et bien établis 309
V. Des réalités plus complexes que les modèles 314
VI. Perspectives de recherche 315
VII. Bibliographie 316

Chapitre 18. L'évolution du concept de feedback, au carrefour des dimensions cognitive, émotionnelle, comportementale et sociale de l'apprentissage 321
Des dimensions inhérentes à l'engagement des apprenants dans le traitement du feedback
Matthieu Hausman, Pascal Detroz et Laurent Leduc

I. Introduction 321
II. Pourquoi fournir des feedbacks aux apprenants ? 323
III. Le paramétrage de l'information et ses limites 324
IV. L'émergence de la « Student Feedback Literacy » 326
V. Le caractère multidimensionnel du traitement des feedbacks 330
VI. Un regard critique sur l'approche contemporaine des feedbacks en éducation 332
VII. Bibliographie 333

Chapitre 19. Culture, pratiques et évolution de l'évaluation en éducation et formation 339
Quarante années de transformation et de recherche au Maroc
Ghizlane Chemsi, Naceur Achtaich, Mohamed Radid et Mohammed Talbi

I. Introduction 339
II. La culture évaluative dans le système éducatif marocain : réformes, acteurs et enjeux 341
III. Pratiques et dispositifs d'évaluation dans les établissements au Maroc 344
IV. La recherche scientifique en évaluation au Maroc : structuration, dynamiques et perspectives 348
V. Conclusion 352
VI. Bibliographie 353

Chapitre 20. Quarante ans d'évolution de l'évaluation au Liban : Enjeux, réformes et perspectives 357
Wassim El-Khatib, Patricia Azoury

I. Introduction 357
II. Méthodologie de l'analyse 358
III. Les premières recherches en évaluation (1980-1990) 360
IV. Institutionnalisation et réforme (1990-2010) 361
V. Innovations et défis contemporains (2010-2020) 364
VI. Vers une politique renouvelée de l'évaluation (2020-2024) 367
VII. Conclusion 369
VIII. Bibliographie 369

Chapitre 21. Du testing adaptatif à l'évaluation intelligente 375
40 ans de tissage entre le numérique et l'évaluation au sein de la Communauté ADMEE
Pierre-François Coen

I. Introduction 375
II. Quelques points de repère 376

III. Les premiers pas du testing adaptatif 378
IV. L'évaluation assistée par ordinateur (EAO) 379
V. Evaluation formative, processus et hybridation de l'évaluation 381
VI. Le boum de la pandémie 383
VII. Les enjeux de l'intelligence artificielle 384
VIII. Constats conclusifs 385
IX. Bibliographie 386

Chapitre 22. Changer notre rapport aux preuves d'apprentissage : questions posées par les IA génératives en évaluation 393
Problématiser l'émergence de l'IA en évaluation des apprentissages en enseignement supérieur
Isabelle Nizet

I. Introduction 393
II. Les piliers de la validité de l'évaluation remis en question 394
II.a. La dimension épistémique des preuves d'apprentissages en évaluation 396
II.b. La dimension axiologique des preuves d'apprentissages en évaluation 397
II.c. La dimension psychopédagogique des preuves d'apprentissage en évaluation 397
III. Problématisation de l'introduction de l'IAg en évaluation 398
III.a. Se positionner sur un continuum d'usages des IA génératives 399
III.b. Identifier le niveau de vulnérabilité des tâches évaluatives à l'IAG 403
III.c. Conserver la plus-value pédagogique des tâches évaluatives 405
IV. Conclusion 407
V. Bibliographie 408

Chapitre 23. 40 ans de crises de valeurs en évaluation 413
Crisologie de nos recherches
Alban Roblez

I. Introduction 413
II. Crise et critique 414
III. La crise (en recherche) dans l'évaluation 416
IV. « Où vont les valeurs ? » Un moment critique dans notre société de l'évaluation 420
V. Et dans 40 ans ? Bref essai de prospective axiologique pour l'évaluation en éducation et en formation 422
VI. Bibliographie 424

Liste des auteurs et autrices 439

Liste des auteurs et autrices

Naceur Achtaich, enseignant-chercheur marocain en mathématiques appliquées, a consacré 40 ans à l'Université Hassan II de Casablanca. Il a dirigé des recherches en analyse numérique et pédagogie, encadré une trentaine de thèses et publié de nombreux travaux. Acteur associatif, il a fondé plusieurs structures pour promouvoir les sciences et l'éducation.

Maîtresse de conférences, membre de l'UMR EFTS à Université Toulouse - Jean Jaurès (France), **Lucie Aussel** est co-responsable du Thème 3 « Organisations et Participation : accompagnement du changement en éducation ». Ses travaux de recherche portent sur les processus de recherches collaboratives en évaluation, l'évaluation de dispositif de formation dans des contextes traversés par un ou plusieurs changements, et la recherche-intervention évaluative.

Patricia Azoury est cheffe du département d'orthopédagogie à l'Université Sainte-Famille Batroun. Consultante en éducation inclusive et membre de réseaux académiques internationaux (ADMEE, ADOQ). Ses recherches portent sur la formation à l'APC, l'adaptation pédagogique et l'inclusion, notamment à travers les stratégies d'intervention et les méthodes d'évaluation inclusives.

Carla Barroso da Costa est professeure à l'Université du Québec à Montréal (UQAM). Ses recherches en mesure et évaluation portent sur la façon dont les acteurs de l'éducation s'autorégulent et s'engagent dans l'apprentissage et l'évaluation, à l'aide de méthodes quantitatives ou mixtes. Depuis 2021, elle dirige la revue *Mesure et Évaluation en Éducation*.

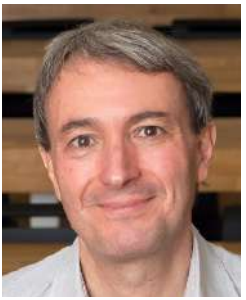
La Pr. **Ariane Baye** – Université de Liège – est spécialiste de l'analyse des politiques éducatives, de l'efficacité et équité des systèmes éducatifs. Elle œuvre à poser les jalons d'une culture de recherche fondée sur des données expérimentales dans son pays et à l'échelle européenne. Pour ce faire, elle réalise des synthèses systématiques des meilleures preuves et met en œuvre des interventions et leur évaluation expérimentale en contexte.

Sébastien Béland est professeur agrégé à la Faculté des sciences de l'éducation de l'Université de Montréal. Il enseigne les méthodes quantitatives et l'évaluation des apprentissages. Ses intérêts de recherche touchent au développement et à l'enseignement des statistiques et de la psychométrie, à l'intégrité académique et à l'évaluation des apprentissages au postsecondaire.

Isabelle Bertolino est chargée de l'évaluation des politiques éducatives au Conseil départemental de la Haute-Garonne. Elle a soutenu une thèse intitulée "Contribution à la définition de la mixité sociale à l'école par l'évaluation du dispositif haut-garonnais". Elle est chercheure partenaire de l'Université Toulouse-Jean Jaurès et cofondatrice de l'association "la Fabrique de la mixité - Laboratoire coopératif et social".

Après avoir été responsable de d'organismes publics (IREM, MAFPEN, IUFM...), président de l'ADMEE-Europe et recteur de l'Académie de Clermont-Ferrand, **Alain Bouvier** a été membre de plusieurs instances chargées d'évaluation de politiques publiques, notamment le Haut conseil de l'éducation pendant 6 ans, puis le Comité parlementaire de suivi de la loi Peillon pendant 5 ans.

Après avoir enseigné dans le premier et second degré, **Dominique Broussal** est aujourd'hui professeur des universités en sciences de l'éducation et de la formation à l'Université de Toulouse et membre de l'Unité mixte de recherche EFTS. Ses travaux mobilisent des démarches de recherche-intervention et interrogent les conditions de cet accompagnement. Le rôle des œuvres dans les dynamiques d'émancipation est au coeur de ses recherches actuelles.



Fabrizio Butera est professeur de psychologie sociale à l'Université de Lausanne. Ses recherches portent sur les processus d'influence sociale, et dans le domaine de l'éducation il s'est intéressé aux processus relationnels et structurels—comme la coopération et la compétition—qui influencent l'apprentissage.

Dominique Casanova est responsable du développement scientifique à la Direction de l'attractivité internationale du Groupe éducatif de la CCI Paris Ile-de-France. Il supervise notamment la qualité métrologique des tests et certifications en langue française du Français des affaires, qu'il représente au sein de l'ADMEE et de l'Association of Language Testers in Europe (ALTE).

Carmen Cavaco. Docteur en sciences de l'éducation/formation des adultes. Professeure et chercheuse à l'Instituto de Educação, Universidade de Lisboa | UIDEF, dans les domaines suivants : formation expérientielle, politiques d'éducation et de formation des adultes, adultes non scolarisés et peu scolarisés, reconnaissance et validation des acquis expérientiels et l'approche biographique.

Christelle Chauffriasse est docteure en Sciences de l'Education et de la Formation. Elle est chercheuse partenaire au sein l'UMR EFTS, inscrite dans le thème « Organisations et participation - Accompagnement du changement en éducation et formation ». Sa thèse, conduite sous la direction de Jean-François Marcel et Lucie Aussel, porte sur la référentialisation quand elle s'inscrit dans une recherche-intervention évaluative.

Ghizlane Chemsi, Docteure en technologie éducative et Maître de Conférences Habilitée à l'Université Hassan II de Casablanca, mène des travaux sur l'évaluation formative, l'ingénierie pédagogique et les technologies éducatives. Elle coordonne des recherches sur l'intelligence artificielle, les learning analytics et l'évaluation des apprentissages, accompagnant la transformation numérique de l'enseignement supérieur.

Pierre-François Coen est professeur à l'Université de Fribourg et responsable de la formation des enseignant·e·s pour les degrés du post-obligatoire. Actif dans le domaine du numérique, ses travaux touchent des problématiques liées à l'intégration du numérique dans l'école en particulier pour développer l'autoévaluation des élèves et la réflexivité des enseignant·e·s.

Dylan Dachet (Ph.D.) est chargé de cours au sein de l'École de Formation des Enseignants de l'Université de Mons. Spécialiste en méta-science et en docimologie, ses travaux portent essentiellement l'évaluation de la qualité de la recherche et sur les biais susceptibles d'en affecter les conclusions. Il a collaboré avec des organismes internationaux tels qu'Education Endowment Foundation et que la Best Evidence Encyclopedia.

Jean-Marie De Ketele est Professeur émérite de l'Université Catholique de Louvain et Professeur invité à l'Institut Catholique de Paris. Il est le Titulaire fondateur de la Chaire UNESCO en Sciences de l'Éducation. Il est Docteur Honoris Causa de l'ICP (Paris) et de l'UCAD (Dakar). Il est rédacteur en chef de la Revue internationale d'éducation de Sèvres et dirige trois collections d'ouvrages scientifiques aux Éditions De Boeck.

Manon Delbreil, ingénieure pédagogique et docteure en Sciences de l'Éducation et de la Formation, est chercheuse associée à l'UMR Éducation Formation Travail Savoirs de l'Université Toulouse Jean-Jaurès. Elle s'intéresse à la pédagogie de l'opportunité et à la co-construction de dispositifs de formation incluant des démarches évaluatives et participatives.

Psychologue et statisticien, **Marc Demeuse** est professeur ordinaire à l'Université de Mons où il dirige l'Institut d'administration scolaire (INAS). Il a contribué à la création des premiers indicateurs européens de la qualité scolaire (2000) et de l'équité dans l'éducation. Ses recherches portent aussi sur des politiques ciblées (éducation prioritaire) et la formation des enseignants. Il a été le premier rédacteur en chef de la revue e-JIREF.

Pascal Detroz est professeur à l'Université de Liège où il dirige l'unité de recherche DidactifEN et assure la responsabilité académique du SMART-IFRES, cellule d'innovation et d'accompagnement en évaluation dans l'enseignement universitaire. Ses travaux portent sur la pédagogie de l'enseignement supérieur, l'évaluation des apprentissages et des dispositifs, la transition secondaire-université et la robustesse dans l'enseignement supérieur.

Annick Fagnant est professeure en Sciences de l'éducation à l'Université de Liège. Ses recherches portent sur l'enseignement/apprentissage des mathématiques dans l'enseignement fondamental, sur l'apprentissage autorégulé et sur l'évaluation en classe. Elle est également impliquée dans plusieurs évaluations internationales telles que TIMSS et TALIS. Elle a été co-fondatrice du réseau EVADIDA et rédactrice en chef du journal e-JIREF.

Céline Girardet est maîtresse d'enseignement et de recherche à l'université de Genève (groupe de recherche « Evaluer, Réguler et Différencier pour apprendre »). Ses recherches portent notamment sur l'évaluation continue pour apprendre, l'évaluation en tant que compétence transversale, le feedback, les émotions dans les processus évaluatifs et l'évolution des croyances et pratiques enseignantes.

Christophe Gremion est professeur à la Haute école fédérale en formation professionnelle HEFP, professeur invité de l'UNamur, chercheur associé du gEvaPP et président de l'ADMEE-Europe. Ses recherches portent sur l'articulation entre évaluation et accompagnement dans les formations à visées professionnalisantes. Il intervient également comme expert externe pour l'évaluation d'établissements scolaires en Suisse et en Belgique.

Agrégé de philosophie, Docteur d'Etat ès Lettres et Sciences Humaines, **Charles Hadji** a exercé les fonctions de Maître de Conférences, de Professeur, aux universités de Lyon, puis de Grenoble. Il est désormais Professeur honoraire (Sciences de l'éducation) de l'Université Grenoble Alpes (UGA). Il a consacré ses principaux travaux de recherche à la philosophie de l'éducation d'une part, et aux problèmes d'apprentissage et d'évaluation d'autre part.

Matthieu Hausman est chargé de cours et de recherche à l'Université de Liège (Belgique). Ses centres d'intérêt sont la gestion pédagogique de l'hétérogénéité des apprenants et l'évaluation des apprentissages. Il est l'auteur d'une thèse de doctorat portant sur les émotions des étudiants dans les situations de feedback en enseignement supérieur. Actuellement co-délégué de la section belge de l'ADMEE-Europe et responsable des Presses de l'ADMEE.

Pauline Hédacq. Puéricultrice au sein du CHU de Toulouse (France), j'ai exercé pendant 15 ans dans les unités d'urgence pédiatrique et de réanimation pédiatrique. J'ai mené mon doctorat au sein de l'unité mixte de recherche Éducation, Formation, Travail, Savoirs de l'université Toulouse Jean Jaurès. Mes travaux de recherche portent sur l'accompagnement du changement par la co-construction de dispositifs d'évaluation ou de formation.

Wassim El-Khatib est professeur à la Faculté de Pédagogie de l'Université Libanaise et expert associé au programme APPRENDRE de l'AUF. Membre des commissions du CRDP chargées de la refonte du curriculum libanais, il s'intéresse à l'évaluation des apprentissages, au développement professionnel des enseignants et à l'innovation pédagogique.

Pascal Lafont, MCF HDR en Sciences de l'Education et de la Formation, UFR SESS-STAPS, Université Paris Est Créteil, LIRTES UR7313. Responsable du parcours « Expertise, Ingénierie et Projets internationaux » de Master en Sciences de l'Éducation. Ses recherches portent sur la Validation des Acquis de l'expérience, les apprentissages non formels et informels, la démocratie participative en milieu académique, l'innovation pédagogique.

Dominique Lafontaine est professeure émérite en Sciences de l'Éducation de l'Université de Liège. Ses recherches ont porté sur l'évaluation, l'efficacité et l'équité des systèmes éducatifs, les politiques éducatives et les enquêtes internationales. Elle a été membre du groupe d'experts en lecture de PISA.

Titulaire d'un doctorat dans le domaine du plan de cours et de la planification pédagogique, **Laurent Leduc** est chargé de cours au sein de l'Institut de Formation et de Recherche en Enseignement Supérieur (IFRES) de l'Université de Liège (ULiège). Ancrés notamment dans le champ de la First-Year Experience, ses travaux sont dédiés au soutien pédagogique et à la formation des enseignant.es en charge des étudiant.es de première année à l'université.

Céline Lepareur est chargée d'enseignement en régulation des apprentissages scolaires à la Haute école pédagogique du canton de Vaud (Lausanne, Suisse). Ses recherches et ses enseignements portent sur les pratiques évaluatives soutenant les apprentissages des élèves et l'apprentissage autorégulé. Elle est actuellement l'une des coordinatrices du réseau « Évaluations et didactique » (EVADIDA) au sein de l'Admée-Europe.

Patrick Mayen - Professeur émérite en sciences de l'éducation Institut Agro Dijon, Président de l'association Recherche et pratiques en didactique professionnelle. Chercheur et intervenant. Travaux de recherche : comment des personnes apprennent-elles et développent-elles des compétences pour et par le travail et par la formation ? Pour l'évaluation : évaluation des compétences, VAE, référentiels et certifications.

Sophie Morlaix est Professeure des universités en sciences de l'éducation et de la formation à l'université Bourgogne Europe. Elle dirige actuellement l'IREDU (Institut de Recherche sur l'éducation). Ses thématiques de recherches portent sur l'équité et l'efficacité des dispositifs et systèmes éducatifs, ainsi que sur la définition, la mesure et les effets des compétences dites non académiques sur les trajectoires scolaires et professionnelles.

Lucie Mottier Lopez est Professeure à la FPSE de l'Université de Genève. Elle dirige le groupe « Evaluer, réguler et différencier pour apprendre dans les systèmes d'enseignement et de formation » (EReD). Ses études actuelles portent sur la voix des élèves dans des évaluations à visée émancipatrice et sur les processus de co-constitution référentielle et de régulation dans une perspective d'évaluation située. Elle a présidé l'ADMEE-Europe de 2009 à 2013.

Isabelle Nizet est professeure titulaire retraitée, associée au département de pédagogie de la faculté d'éducation de l'Université de Sherbrooke, Québec, Canada. Son domaine d'intervention est la construction de savoirs professionnels en évaluation en lien avec les sujets contemporains tels que la décolonisation de l'évaluation, les rapports entre l'intelligence artificielle et l'évaluation ou encore l'évaluation en contexte autochtone.

Françoise Pasche Gossin, professeure à la HEP-BEJUNE (Suisse), dirige le domaine de recherche « Apprentissage et évaluation ». Ses recherches portent sur les questions d'évaluation et de régulation des apprentissages dans différents contextes. Elle s'intéresse également aux enjeux d'enseignement, d'apprentissage et d'évaluation à travers le prisme de l'égalité et de la justice. Elle coordonne la formation à l'enseignement à et par la recherche.

Raphaël Pasquini est Professeur ordinaire en évaluation scolaire à la Haute école pédagogique du canton de Vaud (Lausanne). Ses travaux et ses enseignements portent sur l'évaluation au service des apprentissages, jusqu'à la notation. Il étudie également le développement de la littératie en évaluation des enseignants et leur culture de l'évaluation. Il est actif au sein de réseaux en évaluation aux niveaux régional, national et international.

Cathy Perret est titulaire d'un doctorat en économie de l'éducation et d'une HDR en sciences de l'éducation et de la formation. Ingénieure de recherche au sein de l'IREDU (Institut de recherche sur l'éducation) à l'Université Bourgogne Europe, ses travaux s'intéressent aux politiques et dispositifs d'aide à la réussite et aux problématiques de l'évaluation de l'enseignement supérieur. Elle a été rédactrice en chef de la revue e-JIREF.

Mohamed Radid, Professeur à l'Université Hassan II de Casablanca au Maroc. Ses enseignements et travaux portent sur l'évaluation en éducation et les technologies éducatives. Délégué National de ADMEE Europe. Coorganisateur des colloques ADMEE 2014 et 2020. Coordonnateur du master Ingénierie et Technologies de l'Éducation et de la Formation (2014-2025), Vice-Doyen de la Faculté des Sciences Ben M'Sick Chargé des affaires académiques depuis 2016.

Alban Roblez est maître de conférences enseignant-chercheur à l'Ecole des Sciences de la Société de l'université catholique de Lille. Ses recherches portent sur la fonction épistémologique de la valeur dans les activités d'évaluation et les aspects ordinaires ou alternatifs des pratiques évaluatives. Les contextes d'intervention ou de recherche sont au carrefour de plusieurs disciplines ou champs d'activité, formels, informels et non-formels.

Marc Romainville est professeur émérite de l'Université de Namur. Ses recherches ont porté sur les pratiques étudiantes dans l'enseignement supérieur et sur l'évolution des pratiques enseignantes, notamment en matière d'innovation et d'évaluation des acquis. Plus récemment, il a exploré les apports d'une pédagogie de la métacognition au développement de l'esprit critique face à la désinformation, dans l'ouvrage « À l'école du doute » (PUF).

Mohammed Talbi, Professeur à l'Université Hassan II de Casablanca au Maroc. Ses enseignements et travaux de recherche portent sur l'évaluation en éducation et les technologies éducatives. A été délégué National et membre du CA de ADMEE Europe, Président par intérim de l'Université Hassan II de Casablanca, Doyen de la Faculté des Sciences Ben M'Sik, Conseiller de Messieurs les Ministres de la Santé et des Pêches Maritimes.

Nathalie Younès est professeure des universités en sciences de l'éducation et de la formation à l'Université Clermont-Auvergne, laboratoire ACTé. Elle a été rédactrice européenne de la revue Mesure et Evaluation en Education, présidente de l'ADMEE-Europe, membre de la revue E-JIREF et des Presses de l'ADMEE. Ses recherches portent sur l'évaluation écologique des dispositifs de formation, la culture de l'évaluation et les pratiques d'évaluation.

Cet ouvrage a été publié à l'occasion des 40 ans de l'ADMEE-Europe, soulignant son rôle pionnier dans l'évaluation en éducation. Les autrices et auteurs ont reçu la mission suivante : "un-e futur-e doctorant-e vient vous voir. Iel veut démarrer une thèse dans votre domaine de recherche. Le chapitre pourrait être un résumé de ce que vous lui diriez, expliquant par les 40 dernières années - de l'ADMEE-Europe ou de la recherche en évaluation en général - les lignes de questionnements qui ont conduit aux connaissances actuelles, le-la rendant attentif-ve aux questions vives contemporaines, mais également à celles que vous imaginez pour la prochaine décennie."
La mission a été acceptée par 23 équipes de recherche qui vous livrent ici leurs réflexions.

Christophe Gremion est professeur à la Haute école fédérale en formation professionnelle HEFP, professeur invité de l'UNamur, chercheur associé du gEvaPP et président de l'ADMEE-Europe. Ses recherches portent sur l'articulation entre évaluation et accompagnement dans les formations à visées professionnalisantes. Il intervient également comme expert externe pour l'évaluation d'établissements scolaires en Suisse et en Belgique.

Nathalie Younès est professeure des universités en sciences de l'éducation et de la formation à l'Université Clermont-Auvergne, laboratoire ACTé. Elle a été rédactrice européenne de la revue Mesure et Evaluation en Education, présidente de l'ADMEE-Europe, membre de la revue E-JIREF et des Presses de l'ADMEE. Ses recherches portent sur l'évaluation écologique des dispositifs de formation, la culture de l'évaluation et les pratiques d'évaluation.

9 782970 153689

Presses de l'ADMEE, Faubourg de l'Hôpital 43,
Case postale 556 CH-2002 Neuchâtel SUISSE

www.presses-admee.com