La qualité de l'enseignement : un engagement des établissements, avec les étudiants ?

Laure Endrizzi

▶ **To cite this version:**

Laure Endrizzi. La qualité de l'enseignement : un engagement des établissements, avec les étudiants ?. 2014. ensl-01565660

HAL Id: ensl-01565660
https://ens-lyon.hal.science/ensl-01565660

Submitted on 11 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT FRANÇAIS DE L'ÉDUCATION

n° 93
Juin 2014

Sommaire

● **Page 2 :** Quelle(s) évaluation(s) pour une gouvernance réflexive ? ● **Page 13 :** En France, des établissements plus responsables ? ● **Page 21 :** Évaluation des processus d'enseignement-apprentissage ● **Page 38 :** Bibliographie

Dossier de veille de l'IFÉ

LA QUALITÉ DE L'ENSEIGNEMENT : UN ENGAGEMENT DES ÉTABLISSEMENTS, AVEC LES ÉTUDIANTS ?

Remerciements à Martine Cassette, Isabelle Chênerie, Julien Douady, Michel Le Nir et Cathy Perret pour leur contribution.

Par Laure Endrizzi

Chargée d'étude et de recherche au service Veille et Analyses de l'Institut Français de l'Éducation (IFÉ)

Depuis le Moyen-Âge jusque dans les années 1960, et ce quelles que soient les controverses dont elles ont fait l'objet, les universités étaient, peu ou prou, réservées à une élite. Au XIX[e] siècle, à peine 2 % de la population entrait dans l'enseignement supérieur, aujourd'hui l'Union européenne ambitionne, dans sa stratégie *Europe 2020*, de porter à 40 % la proportion des 30-34 ans en possession d'un diplôme de l'enseignement supérieur, alors que certains pays, tels le Canada, dépassent les 50 %. Cette ambition d'élargir l'accès aux études supérieures, pour autant, n'a de sens que si elle s'accompagne d'une responsabilisation renforcée des établissements concernés, capables de faire preuve de détermination en faveur d'expériences d'enseignement et d'apprentissage réussies et de prendre en compte, d'une manière ou d'une autre, la satisfaction des étudiants dans une démarche d'amélioration continue des formations.

Ce que nous nommons « massification » et rechignons à nommer « démocratisation » – parce que les inégalités n'ont pas disparu, elles se sont déplacées et diversifiées – traduit pourtant un changement d'échelle tout à fait inédit dont nombre d'experts estiment que les universités ont à ce jour insuffisamment pris la mesure. Si elles ne sont plus aussi « faibles » que l'énonçait Georges Felouzis en 2001, leur capacité à planifier le changement de façon systémique reste fragile et certains chantiers n'en sont qu'à leurs balbutiements : l'ouvrage de Donnay et Romainville *Enseigner à l'université : un métier qui s'apprend* (1996) garde toute son actualité aujourd'hui, alors que les enseignants, en France du moins, continuent dans leur grande majorité à reproduire les schémas qu'ils ont connus étudiants : un modèle basé sur l'autoformation, la liberté académique et les implicites, qui sert plus à sélectionner qu'à former.

Toutes les références bibliographiques dans ce Dossier sont accessibles sur notre bibliographie collaborative.

Ce changement de modèle pédagogique n'est cependant pas de la seule responsabilité des enseignants. S'adapter à la diversification et la spécialisation des publics accueillis, miser sur les potentialités offertes par les technologies numériques et s'ajuster aux évolutions de la science et de la société et à celles du marché de l'emploi obligent à un changement culturel profond. L'enseignement n'est pas juste un item dans la liste des missions de l'université, il doit devenir une priorité quotidienne de tous ses membres, y compris les étudiants, car la qualité de l'apprentissage et celle de l'enseignement sont interdépendantes (High Level Group on the Modernisation of Higher Education, 2013). Réduire la distance entre l'université et les étudiants suppose de sortir définitivement du discours de l'inadéquation des publics et de développer une ambition collective explicite pour un enseignement de qualité. Cette « qualité de l'enseignement » ne se laisse pour autant pas définir simplement. Faire réussir des étudiants de première année de droit dans une faculté parisienne prestigieuse ne requiert sans doute pas les mêmes « qualités » que pour faire réussir des étudiants de biologie dans une petite université ou bien encore des étudiants de classes préparatoires littéraires. L'efficacité interne de chaque filière et le contexte institutionnel plus ou moins accompagnant qu'ils découvrent influencent l'entrée dans le métier d'étudiant de ceux qui ne sont plus des héritiers et les incitent peu ou prou à s'engager dans leur formation et à prendre en main leurs apprentissages (Lahire, 1997 ; Coulon, 2005). La relation pédagogique avec l'enseignant et la perception que les étudiants ont de cette relation jouent également un rôle clé dans la réussite de leurs apprentissages (Ramsden, 1988).

L'objectif que se fixe ce dossier est d'examiner les facteurs susceptibles de soutenir un processus d'amélioration continu de l'enseignement en tenant compte, dans la mesure du possible, de cette variance contextuelle. Deux angles sont privilégiés : le premier appréhende la qualité de façon systémique en examinant les modes d'évaluation susceptibles de favoriser une gouvernance réflexive. Le second s'intéresse à l'évaluation des dynamiques d'enseignement-apprentissage et aux utilisations raisonnées et raisonnables de l'évaluation par les étudiants. Un développement spécifique est également consacré à la situation française.

QUELLE(S) ÉVALUATION(S) POUR UNE GOUVERNANCE RÉFLEXIVE ?

Le développement de l'assurance qualité dans les établissements d'enseignement supérieur met en concurrence plusieurs modèles : le modèle étatique centralisé, qui relève davantage d'une logique de contrôle (et donc de sanction) dans un cadre bureaucratique réglementé et un modèle plutôt libéral, qui privilégie une dynamique concurrentielle, des normes d'accréditation et des logiques de classement. Si le curseur tend à se positionner plus souvent du côté du modèle libéral dans les pays anglo-saxons, en Europe globalement, les configurations sont hybrides (Cytermann, 2013).

L'ASSURANCE QUALITÉ : DU PRESTIGE À L'EXCELLENCE ?

Quand la massification oblige à la qualité

Depuis 30 ans, l'image de l'université ● uniquement préoccupée par la production de connaissances scientifiques a considérablement évolué, ses missions se sont diversifiées et la primauté de la recherche sur l'enseignement est fragilisée par l'importance que revêt désormais la relation formation-emploi. Ces changements, mis en évidence au niveau international notamment par les travaux de l'Unesco (réseau GUNI) et de l'OCDE (programme IMHE), s'inscrivent dans un mouvement visant à davantage responsabiliser les établissements, motivé à la fois par la massification et la diversification des publics et par une demande sociale plus explicite pour une meilleure employabilité. Ces pressions pour des établissements plus performants ont largement contribué

● Le terme « université » est employé dans ce dossier comme équivalent d'établissement d'enseignement supérieur. Dans le cas d'une référence explicite à un type d'établissement plutôt qu'un autre (ex. « école d'ingénieurs »), le terme approprié est utilisé.

à actualiser la question de l'évaluation, non seulement relative à leurs activités, mais aussi à leurs objectifs et surtout à leurs résultats. À côté de l'évaluation par les pairs, modèle dominant très largement la culture universitaire, et du contrôle de conformité par les pouvoirs publics, s'est développé un troisième modèle, gestionnaire ou managérial, dans lequel l'université se positionne davantage comme un prestataire de services pour des étudiants-clients. Les tensions entre ces différents référentiels, sensibles partout bien qu'à des intensités variables, touchent à la fois les missions assignées à l'enseignement supérieur, les questions de gouvernance, les modalités de financement et les activités mêmes d'enseignement et de recherche.

Dès le début des années 1990 en Europe, des organismes publics, plus ou moins indépendants, ont reçu pour mission d'évaluer les établissements d'enseignement supérieur pour favoriser le développement d'une culture interne de l'évaluation et l'adoption d'une culture du projet. Diverses formes de contractualisation sont alors apparues entre les pouvoirs publics et les établissements et cette logique contractuelle s'est parfois étendue aux relations entre les directions d'établissement et leurs composantes, ou bien encore aux relations bilatérales ou plurilatérales entre établissements (Dubois, 1998).

C'est dans ce contexte instable que l'« assurance qualité » a pénétré les milieux universitaires, avec la promesse de résoudre les tensions entre les divers modèles ou du moins de restaurer la confiance entre les parties, à l'aide essentiellement de statistiques et d'indicateurs. La création aux Pays-Bas du réseau international INQAAHE, fédérant les agences nationales, est un signe de la propagation de cet intérêt pour la qualité de l'enseignement supérieur, devenue une préoccupation à la fois pour les pouvoirs publics, pour les directions d'établissement et, dans les pays où la relation entre prestataires et usagers domine, pour les étudiants. Signe disent certains d'un **passage d'une logique de prestige, basée sur la seule réputation, à une logique de l'excellence, mondialisée et rationalisée** *via* l'usage d'indicateurs censés produire une mesure fiable de la valeur des institutions (Paradeise & Thoenig, 2011).

À cette « agencification » partout observée en Europe, s'ajoute une multitude d'instruments (classements, prix, labels, concours, indicateurs, réseaux sociaux, etc.) mis en œuvre à des niveaux variés (international, national, local, disciplinaire) par des opérateurs publics et privés (organisations internationales, agences d'accréditation, cabinets de consultance, laboratoires publics ou privés, sociétés du secteur de l'information scientifique, etc.). Ces instruments, par les modèles normatifs qu'ils véhiculent, influent sur l'activité des universités et sur les démarches d'auto-évaluation qu'elles développent.

Les dérives de la qualité à visée normative, potentiellement génératrice d'une débauche d'évaluations et de classements, sont bien connues mais difficiles à éradiquer, à tous les niveaux (voir par exemple Bourdin, 2008 ; Felouzis & Hanhart, 2011 ; Rauhvargers, 2011 ou Hadji, 2012). Quand l'évaluation se pose en finalité plutôt qu'en moyen, quand la compétition, basée sur le contrôle et la comparaison des performances, est érigée en dogme, la qualité, dans sa dimension réflexive et régulative, est dévoyée et sa contribution à la conduite éclairée de l'action est réduite.

La perspective d'une conception raisonnée de la qualité dans l'enseignement supérieur s'inscrit dans la promotion d'une culture de l'auto-évaluation qui combine des mécanismes externes et internes. Autrement dit, pour que la qualité soit « assurée » et que la démarche s'inscrive dans une dynamique positive de changement, il ne suffit pas de produire des indicateurs, qu'ils soient quantitatifs ou qualitatifs. Les **conditions de production et d'utilisation** de ces indicateurs importent et le fait que l'accumulation d'évaluations sans lien les unes avec les autres demeure l'un des principaux écueils n'est pas anodin (Dubois, 1998 ; Bedin, 2009 ; Hadji, 2012 ; Romainville *et al.*, 2013).

Informatique décisionnelle à l'IUT Lumière

De plus en plus sollicité pour produire de l'information quantitative (dossiers d'évaluation pour le ministère, dossiers de renouvellement des flux d'apprentis pour la région…), l'IUT Lumière a souhaité se doter d'un outil d'informatique décisionnelle. Cet outil fait l'objet d'un déploiement progressif sur ses différentes activités (recrutement, suivi de la scolarité étudiante, gestion des relations aux entreprises, suivi des heures, gestion des moyens matériels…). Chaque responsable de diplôme ou de service commun dispose d'une interface lui permettant de récupérer en temps réel indicateurs, diagrammes ou jeux de données à partir des bases développées par l'IUT. Les usagers peuvent participer à la définition des interfaces et au choix des nombreux filtres qui enrichissent ce recueil dynamique des informations. Le temps libéré est mis à profit pour conduire les analyses et les prospectives requises par les dossiers à produire. L'élaboration d'un contrat d'objectifs et de moyens pluriannuel avec l'université Lumière-Lyon 2 et la démarche qualité engagée avec les autres IUT de l'académie de Lyon ont constitué des opportunités pour mobiliser cet outil, tout en lui offrant de nouvelles perspectives de développement.

(Michel Le Nir, université Lumière-Lyon 2, Lyon, France).

Pour les universités européennes qui doivent **apprendre à s'évaluer** (pour renforcer leur autonomie) et **accepter d'être évaluées** (pour préserver leur lien avec les pouvoirs publics, *via* la contractualisation), il s'agit d'une évolution majeure. « *Le modèle, le plus propice à la création d'une dynamique de transformations (performance et qualité) dans chaque université publique met en œuvre une évaluation pluraliste (participative et contradictoire), contextualisée (qui tient compte de l'environnement de l'université), dynamique (qui rapporte l'université à ses objectifs et à son histoire), intégratrice (qui prend en compte l'ensemble des dimensions et des domaines d'activité), et enfin répétée à échéances régulières*. » (Dubois, 1998)

La qualité : aussi une affaire d'enseignement ?

La volonté de se donner les moyens d'offrir un enseignement de qualité, alors que l'offre s'est partout considérablement diversifiée, participe pleinement à cette mouvance. Mais développer la qualité de l'enseignement ● est une entreprise complexe tant l'acte d'enseigner est marqué à la fois par les valeurs de l'établissement dans lequel il se déroule et par celles des acteurs, enseignants et étudiants, qu'il implique. L'arrivée massive des technologies numériques dans le paysage de l'enseignement supérieur contribue largement à la promotion de nouveaux modèles pédagogiques, centrés sur les étudiants ; parallèlement l'internationalisation des formations incite à repenser les curriculums sous l'angle des acquis d'apprentissage ; des centres de ressources sont créés dans les établissements pour accompagner ces transitions et les enseignants sont encouragés à se former et à s'appuyer sur l'expertise de ces centres pour améliorer leurs pratiques. Le défi est d'envergure, car l'activité d'enseignement, traditionnellement individuelle, devient un sujet de débat au sein de divers collectifs (Endrizzi, 2011).

Peu de chercheurs ont à ce jour investi cette question et le lien entre établissement et qualité de l'enseignement est rarement investigué en dehors des travaux comparatifs menés par l'OCDE et l'Union européenne (Fave-Bonnet,

● Dans ce dossier, nous utilisons le terme « qualité des enseignements » pour désigner ce qui se rapporte aux unités d'enseignement (modules, cours). La « qualité de l'enseignement » en revanche concerne à la fois les formations et les unités d'enseignement qui les composent.

2010), alors que les travaux sur l'efficacité de l'enseignement scolaire en revanche, inscrits notamment dans le courant de l'« *educational effectiveness research* » (EER), prospèrent depuis les années 1970 ●. De même « qualité de l'enseignement » ne rime pas spontanément avec « qualité de l'acte d'enseigner ». Bien au contraire, cette préoccupation, assez récente, demeure encore parfois assez éloignée des priorités des pouvoirs publics et des instances dirigeantes, en France en particulier, et exacerbe les tensions inhérentes au statut d'enseignant-chercheur (Endrizzi, 2011).

Il n'est pas rare pourtant que ce soit sous l'impulsion d'éminents chercheurs que les premiers rapprochements entre qualité et enseignement aient été opérés, chercheurs dont les motivations avaient peu à voir avec les théories néolibérales souvent pointées. Ce fut le cas au Royaume-Uni, avec les travaux de Graham Gibbs ●, fondateur de l'ICED (International Consortium for Educational Development) en 1993 et devenu l'un des référents mondiaux en matière de qualité de l'enseignement supérieur, puis avec le rapport Dearing en 1997 ● qui remettait pour la première fois officiellement en cause l'adéquation entre excellence de la recherche et excellence de l'enseignement. Ce fut également le cas en France avec l'engagement de Laurent Schwartz, mathématicien, médaillé Fields, devenu premier président du CNE (Comité national d'évaluation) en 1985 (Mérindol, 2008).

Avant que la qualité des enseignements ne fasse débat, on évaluait plutôt ce qui relève du collectif, à savoir les formations : la qualité était donc envisagée comme « externe », gérée par des opérateurs publics et privés intervenant soit en amont pour délivrer des habilitations, soit en aval pour distinguer telle ou telle formation au moyen de procédures d'accréditation diverses (Fave-Bonnet, 2006). Mais les procédures sont devenues plus complexes, ne serait-ce qu'en raison des ajustements induits par la signature du Processus de Bologne et la mise en œuvre du Cadre européen des certifications ● .

L'évaluation requiert désormais une **implication plus forte des établissements, qui, pour de multiples raisons, ont besoin de faire la preuve de la qualité de leurs formations et donc de se rendre comparables (ou incomparables)**. Il peut s'agir de montrer qu'ils tiennent compte de la demande sociale en faveur d'une meilleure employabilité de leurs diplômés et qu'ils s'inscrivent bien dans une perspective de formation tout au long de la vie. Il peut aussi s'agir de prouver qu'ils sont des opérateurs fiables en termes de formation initiale et continue et qu'ils savent évoluer en misant sur un équilibre entre recherche et enseignement (dans l'optique par exemple d'obtenir des financements). Il peut aussi s'agir d'améliorer leur attractivité, y compris à l'international, en fournissant aux potentiels étudiants des éléments de comparaison avec d'autres établissements, d'autres formations.

Un catalyseur en Europe : le Processus de Bologne

En Europe, c'est la perspective d'un « espace européen de l'enseignement supérieur » (EEES), soutenue par le Processus de Bologne, qui a favorisé la propagation à grande échelle de ces questions de qualité de l'enseignement. Mais les établissements d'Europe n'ont pas attendu Bologne pour se préoccuper de leur performance, au sens large, comme l'ont montré les travaux réalisés dans le cadre du projet EVALUE (Dubois, 1998).

● Une synthèse de l'EER, retraçant son histoire, ses apports théoriques et méthodologiques et ses principaux résultats a été publiée récemment : Reynolds David, Sammons Pam, De Fraine Bieke et al. (2014). « Educational effectiveness research (EER): A state-of-the-art review ». *School Effectiveness and School Improvement,* vol. 25, n° 2, p. 197-230.

● On pourra consulter par exemple Gibbs Graham (1995). « The Relationship between Quality in Research and Quality in Teaching ». *Quality in Higher Education,* vol. 1, n° 2.

● Sir Ron Dearing (1997). *Higher Education in the Learning Society.* Londres : Her Majesty's Stationery Office.

● Les procédures d'accréditation, définies comme l'attribution d'un label à un programme de formation pour sa conformité à des standards spécifiés, nationaux ou internationaux, restent minoritaires en Europe : elles concernent surtout les écoles de management et les écoles d'ingénieurs, et servent parfois dans les pays d'Europe de l'Est à réguler l'offre privée d'enseignement supérieur. Les audits menés par l'ECA (European Consortium for Accreditation) portent essentiellement sur les doubles diplômes ou diplômes conjoints.

À propos du projet EVALUE (1996-1998) piloté par l'université Paris 10

Le projet EVALUE (1996-1998) est une recherche comparative sur l'évaluation et l'auto-évaluation des universités en Europe, financée par le programme de recherche finalisé en socio-économie (TSER) de la Communauté européenne. Il a associé 11 équipes de recherche de 8 pays (Allemagne, Espagne, Finlande, France, Italie, Norvège, Portugal, Royaume-Uni) sous la coordination de Pierre Dubois (université Paris 10, aujourd'hui université Paris-Ouest-Nanterre-La Défense).

Les analyses du consortium, nourries par 31 études de cas d'universités publiques, portent sur toutes les activités des établissements (recherche, enseignement et lien formation-emploi), sur leurs ressources (enseignants-chercheurs, personnels non enseignants) et sur leur gouvernance (modes, moyens, structures). Elles aboutissent à imaginer une université du troisième millénaire dont les conditions d'autonomie et les politiques de contrat lui permettent de s'engager pleinement dans une évaluation pluraliste, contextualisée et dynamique, propice à son évolution (Dubois, 1998).

Au milieu des années 1990, les évaluations externes par des organismes publics tendent déjà à se banaliser et c'est leur contribution au développement de l'évaluation interne qui en justifie majoritairement l'existence. Pour autant, la visée de contrôle n'est pas absente. Les universités, à la fois publiques et autonomes, sont soumises à des pressions financières croissantes, du fait de la progression du nombre d'étudiants, de l'élargissement de l'offre de formation et de la diversification des missions qui leur sont confiées. Elles doivent donc diversifier aussi leurs sources de financement et s'assurer parallèlement de la continuité de l'engagement des pouvoirs publics, nationaux ou régionaux. À cette époque, **la contractualisation est apparue à la fois comme une forme nouvelle de contrôle et un support au développement de l'autonomie universitaire**. De nombreuses initiatives en témoignent, mais elles restent disparates et insuffisamment coordonnées (Dubois, 1998).

Au début des années 2000, dans un contexte où une importance croissante est accordée aux diplômes professionnels (Reverdy, 2014) et aux diplômes de niveau élevé (masters et doctorats), le Processus de Bologne propose une stratégie commune (normalisation des trois cycles d'études et définition des crédits ECTS, obligeant pour la première fois à penser les formations en termes de charge de travail pour les étudiants) et inscrit explicitement la coopération en matière d'assurance qualité au nombre de ses objectifs. La perspective est de faciliter la comparaison des systèmes éducatifs et de garantir la reconnaissance des diplômes à l'intérieur de l'EEES.

C'est dans ce contexte que les ministres des pays signataires ont confié à l'ENQA (European Association for Quality Assurance) la responsabilité de préparer un document de référence guidant les États dans le déploiement de leur propre système d'assurance qualité (ENQA, 2009). Une première version des ESG (*European Standards and Guidelines for Quality Assurance*) a ainsi été adoptée en 2005 lors de la rencontre de Bergen ●. Ce document, contrairement à ce que laisse supposer le terme « *standards* », ne se présente pas comme une liste de critères prescrits, mais propose une démarche basée sur des principes génériques, laissant aux établissements toute latitude pour définir ce qu'ils entendent par « qualité » et sur les procédures à déployer. La démarche s'inscrit ainsi dans un modèle de type « ***fitness***

● La version courante des ESG date de 2009, il s'agit d'une 3e édition, produite par l'ENQA en collaboration avec l'EUA, l'EURASHE et l'ESU. Une traduction en français de l'édition de 2005 est également consultable sur le site de l'ENQA.

for purpose » visant à mesurer l'adéquation des moyens mis en œuvre pour atteindre les objectifs fixés. Elle se décline à trois niveaux (évaluation interne, évaluation externe et fonctionnement des agences nationales) et dans des domaines d'application multiples : des programmes de formation aux disciplines, ou bien des établissements aux unités d'enseignement. Ses visées sont également multiples, depuis la formalisation du rendre-compte et le contrôle des prestations, jusqu'à l'amélioration des enseignements et au final la réussite étudiante, en passant par l'allocation de ressources et l'habilitation des formations (Rege Colet, 2009).

Les multiples rapports produits depuis une dizaine d'années notamment par le BFUG (Bologna Follow-up Group), l'EUA (European University Association), et l'EURASHE (European Association of Institutions in Higher Education) témoignent de l'importance prise par ces démarches qualité et de la diversité des pratiques qu'elles suscitent ●. Les travaux lancés depuis 2008 par la Commission européenne en vue du Cadre européen des certifications, invitant les États membres à redéfinir leur Cadre national en termes d'acquis d'apprentissage (en France, le RNCP, Répertoire national des certifications professionnelles) constituent une incitation supplémentaire à s'engager dans une démarche qualité.

Certains pays ont opté pour la voie constitutionnelle : c'est le cas de la Suisse avec la promulgation en 2011 d'une nouvelle loi fédérale sur « l'encouragement des hautes écoles et la coordination dans le domaine suisse des hautes écoles » (LEHE), pour un système d'enseignement efficace et transparent (Rege Colet, 2009). Plus nombreux sont ceux qui ont installé une agence nationale, comme l'AERES (Agence d'évaluation de la recherche et de l'enseignement supérieur) en France en 2007 ou bien l'AEQES (Agence pour l'évaluation de la qualité de l'enseignement supérieur) en Belgique en 2008. Le registre européen EQAR ● comptait 19 agences labellisées à sa création en 2008, on en dénombrait 32 en 2013, reconnues pour leur conformité aux principes et pratiques promus par l'ENQA (Commission européenne, 2014).

Pour Cytermann (2013), le modèle de qualité promu par Bologne combine **une évaluation par des agences externes et une évaluation interne intégrée**, sans annihiler totalement le rôle de l'État, toujours prépondérant dans la régulation des établissements. Dans le même temps que ces questions d'assurance qualité pénétraient les milieux universitaires, des systèmes de benchmarking se sont développés à l'international, sans que les universités aient prise sur ces initiatives.

LES CLASSEMENTS INTERNATIONAUX : DE L'EXCELLENCE AU PRESTIGE ?

Les universités, fragilisées par une baisse des effectifs étudiants, par la nécessité de restrictions budgétaires et le développement d'une offre privée, ont de plus en plus besoin de se rendre visibles en montrant la qualité de leur formation. Les stratégies de visibilité qu'elles déploient impliquent, peu ou prou, de s'insérer dans des dispositifs de comparaison qui renforcent la concurrence entre établissements et qui modifient leur fonctionnement : il s'agit de s'adapter pour être mieux évalué.

Des classements pour les élites

Les classements internationaux pilotent ainsi en partie les choix de développement des établissements ; la communication des résultats du classement devient plus importante que la qualité réelle des prestations. D'aucuns suggèrent même que les stratégies visant à regrouper les établissements, tels que les PRES (Pôles de recherche et d'enseignement supérieur) en France, ont été motivées par la perspective d'un effet de masse dans les classements internationaux (Vinokur, 2006 ; Fave-Bonnet & Curvale, 2013).

● Le *Report on Progress in Quality Assurance in Higher Education* publié en janvier 2014 par la Commission européenne, dresse la synthèse des dernières études de suivi et analyses réalisées sur ces questions d'assurance qualité. Il s'agit d'une 2e édition, la première datant de 2009 (réf COM(2009) 487 final). Des lectures complémentaires sont disponibles sur le site de l'EEES (ou European Higher Education Area).

● Le registre EQAR (European Quality Assurance Register) accrédite les agences nationales de l'enseignement supérieur qui respectent les principes établis dans les ESG.

Bien évidemment, il est de bon ton de ne pas avouer que la vision portée consiste à copier les institutions les plus prestigieuses et donc à pratiquer ce qu'Annie Vinokur nomme la « *déviation académique* », pour avancer un autre argument, largement exploité par tous ceux qui produisent des classements internationaux ou en diffusent les palmarès : la nécessité de fournir aux usagers des informations objectivées pour choisir leur établissement.

Mais l'argument est fallacieux : si quelques travaux sur la mobilité internationale mentionnent cette fonction décisionnelle, ce type de stratégie estudiantine semble jouer davantage sur le choix du pays que sur celui de l'établissement (Varghese, 2008) et les universités qui en bénéficient sont celles qui disposent déjà d'un fort capital réputationnel (voir par exemple ACA, 2006) ●. Aucun classement international, en tout état de cause, n'offre la granularité requise pour comparer des programmes de formation et les méthodologies déployées permettent difficilement d'apprécier la qualité de l'enseignement, sauf à considérer qu'une recherche excellente va de pair avec un enseignement de qualité (Bourdin, 2008 ; Rauhvargers, 2011, 2013) ●.

Ainsi, le palmarès de l'ARWU (*Academic ranking of world universities*), ou « classement de Shangaï », basé sur des indicateurs quantitatifs de productivité scientifique, ne fait que confirmer, depuis 2003, ce que tout le monde sait déjà, à savoir l'excellence des grosses universités américaines à produire des prix Nobel (ou des médaillés Fields) et à publier dans les revues *Nature* ou *Science*. L'évolution récente vers des classements disciplinaires parallèles ne modifie pas la liste des 500 universités classées, elle permet juste de pondérer leurs profils disciplinaires.

L'examen du très médiatisé THE World university ranking produit par le *Times Higher Education* depuis 2004, ne s'avère guère plus probant, au regard de la qualité de l'enseignement. Supposé viser un public large peu préoccupé par la bibliométrie, il présente un palmarès relativement convergent avec celui de Shangaï… parce que la méthodologie employée n'est pas fondamentalement différente, malgré l'inclusion de critères relatifs à l'enseignement (Rauhvargers, 2013) ●.

Ainsi, pour reprendre les clés de lecture proposées par Paradeise et Thoenig (2011), tout se passe comme si les classements internationaux lancés au milieu des années 2000, en cherchant à mesurer l'excellence, ne parvenaient qu'à sélectionner quelques centaines d'établissements d'élite dans le monde pour leur conformité à certains canons scientifiques peu représentatifs de la diversité des courants de recherche et ignorants des autres missions de l'enseignement supérieur (De Ketele, 2013).

Mais ce benchmarking affecte l'enseignement supérieur bien au-delà du périmètre constitué par les 700 à 1 200 établissements d'élite éligibles. Alors que la demande pour plus de repères factuels s'accroît, tant du côté des citoyens que de celui des pouvoirs publics, certaines universités considèrent que leur mission d'enseignement peut devenir une caractéristique distinctive et revendiquent le fait que leurs atouts puissent être connus (Hénard, 2010). Ce qui implique de réviser complètement les méthodes de collecte des premières enquêtes et de favoriser une diversification des instruments et des dimensions considérés, notamment dans la perspective d'offrir à ceux qui les utilisent la possibilité de classer les établissements en fonction des critères qui ont du sens pour eux (Rauhvargers, 2013).

Vers des classements multidimensionnels

C'est en 2008 que la Commission européenne a entrepris de financer la conception d'un outil multicritères adapté à d'autres types d'établissements que les universités d'élite, nommé U-Multirank (UMR). L'enjeu est de parvenir à catégoriser les établissements d'enseignement supérieur et à les différen-

L'étude de Varghese (2008) montre que les mouvements migratoires vont des pays où le système d'enseignement supérieur est peu développé vers les pays où les établissements bénéficient d'une meilleure réputation, dans le cadre de mobilités intra-régionales en pleine croissance. Celle de l'ACA (Academic Cooperation Association) met en évidence l'attractivité des universités américaines auprès des étudiants originaires d'Asie, car elles sont considérées comme plus prestigieuses et plus innovantes que celles des pays européens.

La donne est sans doute un peu différente aux États-Unis, alors que le classement national America's Best Colleges, réalisé par US News and World Report, exerce une influence notable sur les choix d'orientation des néo-étudiants. Pour autant, le palmarès est assujetti à une logique réputationnelle construite sur des critères liés à la performance en recherche (montant des subventions de recherche, nombre de doctorats, etc.) et au niveau de sélection à l'entrée, sans lien avec la qualité de l'enseignement (Gibbs, 2010).

D'une part les mêmes critères relatifs à la recherche sont mobilisés (taux de citation extraits du Web of Science de Thomson Reuters, depuis 2011) ; d'autre part plus de 30 % des résultats proviennent d'une enquête de réputation auprès d'enseignants-chercheurs et d'employeurs.

cier par des points forts et des points faibles qui touchent à l'ensemble de leurs activités. Après une étude de faisabilité (van Vught & Ziegele, 2012), la plateforme, largement inspirée des travaux néerlandais du CHEPS (U-map) et du système allemand CHE University Ranking, a été inaugurée en mai 2014 ● ; elle référence 850 universités (dont 62 % sont européennes et une trentaine sont françaises) sur la base d'informations qu'elles ont elles-mêmes fournies ou qui ont été collectées dans des bases de données publiques.

Le classement multidimensionnel intègre des indicateurs relatifs aux cinq critères clés suivants : performances en matière de recherche ; qualité de l'enseignement et de l'apprentissage ; ouverture sur l'international ; réussite en matière de transfert de connaissances ; actions et engagement au niveau régional. L'activation de tel ou tel critère permet de générer des classements à la volée et de rendre ainsi « visibles » des établissements qui n'apparaissent pas dans les autres classements internationaux.

UMR développe également une approche disciplinaire et prend en compte, quand elles sont disponibles, les évaluations faites par les étudiants. Les disciplines concernées pour l'instant sont les études commerciales, le génie électrique, le génie mécanique et la physique ; en 2015 s'y ajouteront la psychologie, l'informatique et la médecine ●.

En France, c'est l'OST (Observatoire des sciences et techniques), mandaté par la CPU (Conférence des présidents d'université), la CGE (Conférence des grandes écoles) et la CDEFI (Conférence des directeurs des écoles d'ingénieurs françaises), qui se charge du déploiement d'UMR, *via* la mise en œuvre d'une interface interopérable. Le CERES (outil de caractérisation des établissements de recherche et d'enseignement supérieur) repose sur les mêmes principes qu'UMR (cf. dossier de presse de l'OST et de la CPU, mars 2013).

Une bien belle idée. Reste à s'assurer que les données sources sont suffisamment homogènes pour que les établissements soient effectivement comparables… Le modèle actuellement privilégié d'un dispositif fédératif international, basé sur la production de classements nationaux incluant un tronc commun compatible avec UMR, ne fait pas qu'assurer un avenir radieux à tous ceux qui veulent devenir « responsable qualité »… Il a aussi un coût.

L'AUTO-ÉVALUATION : UNE APPROCHE MULTIDIMENSIONNELLE

Si la qualité d'un établissement ne se rapporte plus exclusivement à ses performances en matière de recherche, de nouvelles méthodes doivent être élaborées pour collecter d'autres types de preuves : il s'agit alors d'adopter une approche multidimensionnelle qui inclut aussi la gouvernance et l'enseignement (au sens d'acte d'enseigner) et qui vise à la fois l'efficience (rationalisation des moyens, retour sur investissement…) et l'efficacité (amélioration des enseignements et des apprentissages, réussite étudiante…), voir Fave-Bonnet (2006).

Vers une évaluation interne qui doit « faire système »

La qualité de l'enseignement résulte ainsi d'une combinaison de facteurs relevant de trois niveaux interdépendants : le niveau de l'établissement (élaboration et mise en œuvre des politiques institutionnelles, déploiement des systèmes internes d'assurance qualité…) ; celui des formations (actions visant à concevoir et actualiser les programmes, à définir les modalités pédagogiques…) ; et le niveau des individus (actions visant à soutenir les enseignants dans leur mission et à aider les étudiants à réussir…).

Le défi est donc d'envergure pour les établissements d'enseignement supérieur qui sont tenus de trouver un équilibre entre les contraintes externes et les potentiels internes, en s'aventurant simultanément sur plusieurs tableaux : développer une culture qualité, construire une offre de formation de qualité et soutenir l'enseignement et l'apprentissage (Hénard, 2010).

● UMR est l'un des rares classements internationaux qui ne résulte pas d'une initiative locale. Il a été conçu par un consortium indépendant placé sous la direction du Centrum für Hochschulentwicklung (CHE) en Allemagne et du Center for Higher Education Policy Studies (CHEPS) aux Pays-Bas. L'un propose une évaluation qui s'appuie sur les résultats des évaluations par les étudiants, l'autre favorise la comparaison entre des profils d'établissement proches.

● Les établissements sont évalués sur la base d'une trentaine d'indicateurs associés à cinq niveaux de performance allant de A (très bon) à E (faible). Les premiers résultats montrent que si plus de 95 % des établissements reçoivent un A sur au moins un point, seuls 12 % d'entre eux cumulent plus de 10 fois la note maximale.

Les changements induits par cette conception dynamique de la qualité sont loin d'être anodins. D'une part, la mise en cohérence de tous ces composants est perçue désormais comme nécessaire, soutenue par une **évaluation qui « fait système »** et qui participe à une gouvernance rationalisée. D'autre part, l'évaluation des enseignements et des apprentissages, dans toutes les ramifications qu'elle est susceptible d'encourager, est à mettre au crédit de la qualité d'un établissement d'enseignement supérieur (Fave-Bonnet, 2010).

Les travaux du High Level Group on the Modernisation of Higher Education (2013) pour la Commission européenne abondent dans cette direction d'une **responsabilisation collective** au niveau de l'établissement, tout en insistant sur le rôle d'aiguillon des politiques publiques. Hénard (2010) distingue également quatre types d'acteurs internes de premier plan :

- les équipes dirigeantes et instances décisionnaires de l'établissement qui sont souvent à l'origine du projet et qui jouent un rôle clé dans la définition de ses objectifs, dans les moyens à mobiliser et dans la diffusion et appropriation de cette culture qualité ;
- les enseignants, qui doivent être convaincus de disposer d'instruments de mesure adéquats et certains de bénéficier d'un soutien également adéquat, qui idéalement y voient une opportunité pour améliorer leur enseignement tout en situant mieux leur action individuelle à l'échelle de l'établissement ;
- les étudiants, premiers bénéficiaires de cette démarche qualité, qui doivent être partie prenante en collaborant avec les dirigeants et les enseignants pour définir le périmètre de ces évaluations, soit au titre de leur position d'élus dans telle ou telle instance, soit au nom de l'association dont ils se revendiquent ;
- les cellules qualité ou structures d'appui avec un mandat explicite qui peuvent assumer une fonction pivot tout à fait essentielle à différents stades du projet : en amont pour concevoir les instruments ou pour soutenir méthodologiquement et/ou techniquement les facultés, et en aval pour interpréter les résultats et accompagner la remédiation.

Le fait d'installer une structure (ou une fonction) permanente dans l'établissement renforce la mise en cohérence et la continuité de la démarche. La participation des différents acteurs internes au processus d'évaluation peut garantir une meilleure acceptation des résultats et favoriser les effets d'apprentissage à la fois individuels et collectifs (Dubois, 1998). Il ne s'agit pas de faire nécessairement converger toutes les opinions, mais de réduire les tensions entre les différents points de vue (Lanarès, 2009). Pour Hadji (2012), l'évaluation qui vise une **amélioration des pratiques individuelles et collectives** doit faire sens pour tous, et pas seulement pour les commanditaires. Chacun doit ainsi comprendre ses forces et ses faiblesses en tant que membre d'un collectif.

De nombreuses initiatives ont d'ores et déjà été prises à l'un ou l'autre des trois niveaux précités (établissement, formations, individus) pour améliorer la qualité de l'enseignement : créer un centre de soutien à l'enseignement, proposer des activités de développement professionnel (formations par exemple), récompenser les enseignants particulièrement investis, constituer un fonds de soutien à l'innovation pédagogique, recruter des personnels sur critères pédagogiques, miser sur un management plus « collectif », investir dans des équipements facilitateurs (informatique, documentation…), mettre en place des services d'appui performants (tutorat, orientation et insertion…), miser sur des dispositifs d'évaluation par les étudiants, encourager l'auto-évaluation et la co-évaluation chez les enseignants, déployer des outils décisionnels pour analyser les données quantitatives collectées, etc.

Les travaux récents de l'OCDE (Hénard, 2010) et de l'Union européenne (Commission européenne, 2014) confirment que les initiatives se sont multipliées au sein des établissements d'enseignement supérieur un peu partout dans le monde, et montrent qu'elles traduisent **une vision de la qualité à la fois polymorphe et dynamique, mais insuffisamment stratégique**, du fait justement d'un fort émiettement. Ces initiatives demeurent très majoritairement empiriques et s'appuient très modestement sur les ESG de l'ENQA (2009). De ce fait, leur consolidation dans le cadre d'une politique institutionnelle reste un enjeu fort à court et moyen termes pour les universités (Hénard & Roseveare, 2012 ; High Level Group on the Modernisation of Higher Education, 2013).

À propos du guide *Fostering quality in higher education: Policies and practices (2012)*
Ce guide, édité par l'OCDE dans le cadre du projet « Qualité de l'enseignement dans le supérieur », vise à aider les présidents d'université et les responsables académiques à élaborer des politiques en matière de qualité des formations et des enseignements. Il comprend des scénarios, des cadres de réflexion et des exemples de pratiques. Pour les auteurs, sept leviers sont à actionner : sensibilisation des différentes parties, développement professionnel des enseignants, engagement des étudiants, dynamique du changement et leadership partagé, alignement des multiples initiatives en faveur de la qualité, investissement dans l'innovation et évaluation des impacts (Hénard & Roseveare, 2012).

La promotion de cette vision plus stratégique passe par une **démarche objectivée d'auto-évaluation qui combine évaluation externe et évaluation interne, la première contribuant au développement de la seconde**. Cette démarche conduit par ailleurs à l'émergence d'un nouveau type d'acteur : les pairs-évaluateurs, qui en tant qu'enseignants-chercheurs, doivent parvenir à résoudre les conflits potentiels entre leur fonction d'évaluateur et leur identité professionnelle pour jouer un véritable rôle de régulateur (King *et al.*, 2007), et dont l'action peut s'interpréter à la lumière de la constitution d'une nouvelle élite, battant alors en brèche le discours commun sur l'affaiblissement de la profession académique (Musselin, 2013).

La révision en cours des ESG doit permettre d'en améliorer la clarté, l'applicabilité et l'utilité dans cette perspective systémique (Commission européenne, 2014). La version provisoire, rendue publique en mars 2014, établit une relation plus explicite entre le processus d'enseignement-apprentissage et l'assurance qualité interne ; en intégrant les développements majeurs intervenus depuis 2005 (définition des cadres de certification et de reconnaissance des diplômes, promotion des acquis d'apprentissage), elle a vocation à offrir un cadre renouvelé fondé sur une conception plus pédocentrée de l'enseignement.

Quelles dimensions pour la qualité de l'enseignement ?

Les dispositifs d'auto-évaluation traduisent des conceptions différenciées de ce qu'« enseignement » et « qualité » veulent dire. Les objectifs priorisés et les dimensions estimées légitimes pour mesurer cette qualité sont à comprendre comme autant de marqueurs de la « culture d'établissement » à un moment donné, qu'ils émanent essentiellement des instances dirigeantes ou d'un collectif plus participatif. Les partenariats avec les milieux professionnels ou avec les établissements étrangers, la cohérence de l'offre de formation à une échelle locale ou régionale, la diversité des publics accueillis (notamment la proportion d'étudiants internationaux), les investissements en

matière d'environnement numérique ou de ressources humaines ou bien encore les taux d'insertion des diplômés peuvent par exemple être considérés, explicitement ou implicitement, comme des critères plus ou moins cruciaux.

Sur un plan théorique, il n'existe pas d'unanimité sur la définition de ce qui caractérise un enseignement de qualité, ni sur la façon adéquate de le mesurer. Pour certains, c'est la raison pour laquelle il est indispensable d'opter pour une approche concertée à ce stade de la démarche (Romainville, 2009 ; Younès & Romainville, 2012). Dans la conception appréciative de l'évaluation promue par Hadji (2012), l'évaluation résulte d'une confrontation entre les attentes exprimées sous forme de « critères » et des aspects significatifs du réel traduits en « indicateurs » pour mesurer l'adéquation entre ce qu'on est légitimement en droit d'attendre et la réalité évaluée. Autrement dit, **c'est le réalisme des attentes qui doit guider ce sur quoi l'évaluation porte**.

Évaluation des enseignement à Lille 1 : une démarche intégrée
Commencée dès 1998, l'évaluation des enseignements est devenue systématique en 2004 (passage au LMD). Elle prend deux formes :
- une évaluation des enseignements, matière par matière, sur un niveau d'études (par exemple les étudiants de 3e année de licence) ; les questions programmées en ligne sur les matières sont les mêmes pour l'ensemble des parcours (sciences, économie et sciences humaines) ;
- une évaluation d'un parcours disciplinaire sur plusieurs niveaux d'études (par exemple les étudiants de 1re et 2e années de master de biologie) sur les activités pédagogiques autres que les enseignements (conférences, mémoire, stages) ou sur la structuration et les options des parcours des étudiants.

L'Observatoire des formations et de l'insertion professionnelle (OFIP) a en charge la réalisation des évaluations définies par le CEVU (préparation des questionnaires, analyse des résultats et production des documents) et implique les responsables pédagogiques dans la préparation et la mise en œuvre de l'évaluation.
Une nouvelle étape vient d'être réalisée avec la création du Centre d'accompagnement des pratiques enseignantes.
(Martine Cassette, université Lille 1, Lille, France).

Pour Gibbs (2010), le modèle des « 3P » (*presage, process, product*) défini par Biggs ● en 1993 permet de penser les dimensions qui peuvent raisonnablement être comparées dans le cadre d'une démarche qualité. Il combine :
- des éléments contextuels propres à l'établissement mais extérieurs à la situation d'enseignement-apprentissage (ressources disponibles, niveau de sélection à l'entrée, qualités du public accueilli, qualités des enseignants, nature et poids de la recherche, etc.) ;
- des critères relatifs au processus même d'enseignement-apprentissage (conséquences de la taille des classes, niveau d'effort et d'engagement des étudiants, nature et quantité du feedback reçu sur leur travail, etc.) ;
- et d'autres enfin correspondant aux résultats de cet enseignement-apprentissage (acquis d'apprentissage, diplomation, rétention ou maintien en formation, employabilité, etc.), avec l'idée de privilégier le « gain » relatif à la formation, plutôt que la performance.

Ces combinaisons, quelles que soient les méthodes utilisées pour les obtenir (théories de l'apprentissage, recherches empiriques, analyses logiques, etc.) s'incarnent

● Biggs J. (1993). « From theory to practice: A cognitive systems approach ». *Higher Education Research and Development*, vol. 12, n° 1, p. 73-85. Ce modèle est assez proche de celui proposé par A. Astin aux États-Unis et utilisé dans plusieurs enquêtes à grande échelle : le modèle « I-E-O » (*Input-Environment-Output*).

dans des modélisations multiples qui incluent un nombre extrêmement élevé de dimensions et d'items. Selon Gibbs (2010), **ce sont les variables relatives au processus**, permettant de s'assurer que certaines pratiques pédagogiques sont bien mises en œuvre, **qui prédisent le mieux la qualité des apprentissages**. Ce sont aussi en général les variables les moins bien documentées et les plus controversées, c'est pourquoi elles doivent concentrer les efforts d'objectivation. Les critères nécessaires (et suffisants) pour évaluer ce processus d'enseignement-apprentissage regroupent selon lui la taille du groupe classe, le niveau d'effort et d'engagement des étudiants, le profil des enseignants (expérience et formation) et la qualité et quantité du feedback reçu par les étudiants sur leur travail.

C'est aussi la position adoptée par Abrami *et al.* (2007) qui proposent de privilégier l'impact du processus sur le produit « *defined as the instructor activities which occur both before (preparatory) and during (delivery) teaching which produce positive changes in students in relevant academic domains including the cognitive, affective, and occasionally the psychomotor ones* ». Pour Bernard (2011), les dimensions déterminantes, caractéristiques des situations d'enseignement, sont la planification de cours, la prestation en cours, l'évaluation des apprentissages et la connaissance de la matière enseignée. Si les trois premières ont vocation à être évaluées par les étudiants, l'auteur signale qu'un débat subsiste sur la quatrième, certains chercheurs estimant que la connaissance disciplinaire relève davantage de la responsabilité des pairs.

Pour les étudiants interrogés par l'ESU (European Students' Union, voir Galan Palomares *et al.*, 2013), les points à examiner ne relèvent pas en revanche exclusivement de ce qui se passe à l'échelle d'un cours donné. Quatre dimensions sont estimées déterminantes : un curriculum équilibré entre connaissances et compétences, des modalités d'enseignement qui les rendent acteurs, un environnement d'apprentissage qui réponde à leurs besoins et des équipements et services qui soutiennent efficacement leur vie d'étudiant.

À travers ces quelques exemples, on voit bien que la qualité est difficilement réductible à quelques dimensions génériques et que des **tensions entre le multidimensionnel et le contextuel** semblent inévitables. Des tensions qui risquent d'être encore plus aiguës quand il s'agit de déterminer les critères et les indicateurs propres à chacune d'elle, dans le respect de la complexité des dynamiques d'enseignement et d'apprentissage et le cas échéant en cohérence avec les spécificités disciplinaires...

EN FRANCE, DES ÉTABLISSEMENTS PLUS RESPONSABLES ?

UNE IMPULSION NATIONALE AVEC LA CONTRACTUALISATION

En France, c'est dans les années 1980 que l'évaluation entre dans les établissements d'enseignement supérieur (Thélot, 1993), avec notamment l'évaluation des productions scientifiques au sein des sections disciplinaires du CNU (Conseil national des universités) et le contrôle des maquettes de diplômes par le ministère. La création du CNE (Comité national d'évaluation) en 1985, dans le cadre de la loi Savary, ne remet pas en question ces formes d'évaluation : les audits réalisés sont clairement mis au service du développement d'une culture du projet dans les établissements d'enseignement supérieur.

C'est donc la mise en place de la politique contractuelle dès 1987 qui conduit à des rapprochements pérennes entre évaluation et enseignement supérieur, même si la normalisation de ces rapprochements, soumise à l'alternance du pouvoir politique, reste fragile aujourd'hui. D'abord appliquée aux contrats pluriannuels de recherche au début des années 1980, la contractualisation, étendue aux universités, modifie en profondeur la nature de leurs relations avec l'État et contribue à légitimer le principe de l'évaluation.

« *La démarche contractuelle est fondée sur le fait que l'établissement se dote d'un projet, base de la négociation avec le ministère, par lequel cet établissement se fixe des objectifs et décrit les moyens par lesquels il compte les atteindre. Et le ministère évalue ce projet, à la fois pour savoir s'il est de qualité et s'il rencontre les préoccupations de la politique nationale de l'enseignement supérieur et de la recherche. […] Si l'on prend au sérieux l'autonomie et la démarche contractuelle, l'évaluation ne peut se réduire à mesurer la conformité d'un établissement, ou de ses composantes, aux indicateurs de performance que se donne l'État. Autrement dit, l'évaluation d'un établissement autonome ne peut faire l'impasse sur une connaissance, et une analyse, de sa stratégie. Ce qui a non seulement un impact sur les critères choisis pour l'évaluation, mais aussi sur les méthodes suivies* » (Mérindol, 2008).

Dans ce contexte, les audits du CNE accordent un intérêt pondéré aux statistiques. L'objectif poursuivi est de fournir des pistes diagnostiques aux établissements dans l'ensemble de leurs activités : formation initiale et continue, recherche scientifique et valorisation, diffusion de la culture scientifique et coopération internationale. Ils servent d'appui aux pouvoirs publics lors de l'examen du projet d'établissement mais ne conditionnent en rien l'allocation des ressources ● .

Dans cette dynamique en faveur de la contractualisation, est également créé en 1990, en complément du CNE, un autre organisme indépendant, l'OST, chargé

Le CNE pratiquera ses audits jusqu'en 2005, date à laquelle la création de la future agence d'évaluation de la recherche et de l'enseignement supérieur est déjà décidée. Le site du CNE est toujours actif ; on y retrouve l'ensemble des rapports publics par établissement et par thème, les rapports d'activité annuels adressés au Président de la République et une présentation détaillée de la méthodologie employée.

Dans la phase préalable d'auto-évaluation, les établissements peuvent s'appuyer sur le ***Livre des références*** (LDR), outil adaptable conçu pour les aider à identifier leurs forces et leurs faiblesses en termes de formation, de recherche et de management. En voici un extrait (version de novembre 2003, p. 14) en lien avec la qualité des formations :

Référence A.II.3 – L'offre de formation fait l'objet d'une évaluation régulière.

- Critère 1 : Les enseignements sont évalués régulièrement.
- Critère 2 : Les étudiants sont associés aux évaluations des enseignements et des formations.
- Critère 3 : Les milieux professionnels participent aux évaluations.
- Critère 4 : Les résultats des examens et des concours sont analysés et diffusés.
- Critère 5 : Des commissions s'appuient sur les résultats des évaluations pour faire évoluer l'offre de formation et les enseignements.
- Critère 6 : L'insertion professionnelle des diplômés fait l'objet d'un suivi permanent qui est publié.

Référence A.II.4 – L'établissement a une politique d'amélioration de la pédagogie.

- Critère 1 : Il existe un service ressource, en liaison avec le CEVU, qui favorise la réflexion pédagogique.
- Critère 2 : Il existe des dispositifs de soutien à l'expérimentation pédagogique.
- Critère 3 : L'engagement pédagogique des enseignants-chercheurs est pris en compte lors de l'examen des promotions locales.
- Critère 4 : Des formations sont proposées aux enseignants.
- Critère 5 : L'établissement organise la prise de responsabilités pédagogiques par les enseignants.
- Critère 6 : L'établissement a mis en place des moyens de suivi des formations.
- Critère 7 : Les différents personnels concernés sont associés à la mise en œuvre de l'offre de formation.

Une *Petite histoire de l'évaluation de l'enseignement à l'université*, publiée par Isabelle Chênerie en 2005, retrace les grandes étapes de la progression de la culture qualité en France.

« *La différence entre la mesure de l'effort individuel et celle du résultat presque toujours collectif de cet effort constitue un des plus difficiles problèmes de l'évaluation. Le problème existe dans toutes les activités humaines. Mais il est considérablement aggravé pour l'activité universitaire du fait des spécificités de l'organisation universitaire : chaque enseignant agit comme un travailleur indépendant ou plutôt comme une artisan autonome directement en charge de sa contribution. L'organisation spontanée du travail n'est pas l'équipe intégrée mais la guilde de protection des artisans. Or le seul résultat qui compte et qui soit réellement mesurable, c'est ce que les étudiants tirent de l'ensemble de leurs prestations, c'est-à-dire du travail collectif de tous les enseignants en charge du programme qu'ils suivent.*
Une bonne évaluation ne peut donc porter directement sur les performances individuelles des enseignants. Elle doit viser d'abord le résultat collectif et avoir pour objectif de faire prendre conscience à chaque enseignant de sa contribution à ce collectif » (Crozier, 1990).

de la production d'indicateurs relatifs aux activités de recherche des établissements.
De nouvelles polémiques vont alors voir le jour avec la publication de plusieurs rapports commandités par le ministère. Tous alertent sur le fait que les évaluations déployées sont de mauvaise qualité et qu'il convient de les améliorer en misant sur plus de transparence et d'objectivité. Tous plaident peu ou prou pour une qualité pédagogique renforcée et pour une association des étudiants à la démarche d'auto-évaluation :

- le rapport de Michel Crozier (1990), *L'évaluation des performances pédagogiques des établissements d'enseignement supérieur*, préconise une mise en œuvre de l'évaluation par les établissements eux-mêmes et s'inscrit dans une approche régulative, qui viserait notamment à sensibiliser les enseignants et leurs dirigeants à une appréhension collective de l'activité d'enseignement. Il plaide par exemple pour une évaluation des formations qui aiguillonne la notion d'« équipe » ;
- le rapport d'Alain Lancelot (1995), *L'évaluation pédagogique des universitaires*, développe une approche plus normative et donc plus prescriptive. Il distingue des critères relatifs à la cohérence des formations et à la qualité pédagogique. Sur ce point précisément, il propose des critères assez formels (*a minima* ponctualité, assiduité, respect des heures de services statutaires, régularité et rapidité avec laquelle les copies sont corrigées) et suggère que l'évaluation par les étudiants ne commence qu'au second cycle ;
- le rapport de Maurice Gomel (1996), *L'évaluation de la qualité d'un enseignement supérieur. Évaluer qui ou quoi et comment ?*, reste dans la tonalité prescriptive du rapport Lancelot, tout en proposant une approche moins étroite. Il distingue l'évaluation de l'activité pédagogique individualisée des universitaires, l'évaluation de la pertinence des enseignements universitaires (finalités, contenus…) et l'évaluation de la qualité pédagogique (méthodes, techniques et stratégies d'enseignement) ;
- le rapport Dejean (2002), *L'évaluation de l'enseignement dans les universités françaises*, a nourri les conclusions de l'avis publié par le Haut conseil de l'évaluation de l'école la même année et inspiré l'arrêté Lang du 23 avril 2002 abrogé depuis. Il observe que l'évaluation de l'enseignement reste rare et peu soutenue dans les universités françaises, que ses enjeux sont mal perçus par l'ensemble des acteurs et que sa mise en œuvre rencontre des obstacles importants, notamment du fait d'une absence de suivi au niveau central. Le HCEE conclut que l'État doit se positionner plus fermement et opter en faveur de mesures concrètes pour d'une part encourager les expériences positives et en faire des points d'appui et pour d'autre part donner du sens à ces expériences, en arbitrant sur la place de l'évaluation de l'enseignement dans les procédures d'habilitation et de contractualisation, dans les évaluations externes (menées à l'époque par le CNE) et dans la carrière des enseignants-chercheurs.

Ces questions d'évaluation de la qualité de l'enseignement, notamment par les étudiants, ont en effet du mal à se développer en France, encore aujourd'hui. De nombreux obstacles liés à la fois à l'évaluation elle-même et à la non-reconnaissance des activités d'enseignement sont susceptibles d'en freiner le déploiement, tant à un niveau individuel que collectif (voire corporatiste). Dejean en dresse en 2002 une analyse qui n'a rien perdu de son actualité 12 ans après. Le fait que la loi Savary de 1984 prône « *la pleine indépendance* » et « *l'entière liberté d'expression* » des enseignants et chercheurs n'est pas sans incidence sur cette non-progression. Plus fondamentalement, elle donne une assise juridique au statut d'exception dont les universitaires français se réclament et constitue une belle illustration de ce que D'Iribarne (1989) nomme la « *logique de l'honneur* » : tout comme dans les systèmes de castes, ce sont alors les droits et devoirs

que se donne le groupe d'appartenance, en fonction du rang qu'il occupe par rapport à d'autres groupes, qui en dictent les agissements, contrairement à la logique de la vertu qui incite à respecter des lois qui s'appliquent à tous (Dejean, 2002).

En tout état de cause, la qualité de l'enseignement ne semble pas encore un enjeu en voie de normalisation. Les audits du CNE, à cet égard, n'ont pas fait levier ou du moins les 20 ans qu'ont duré ses actions n'ont pas suffi au Comité pour pénétrer les cultures professionnelles. Bien évidemment les établissements ont investi dans la conception de dispositifs qui les aident à acquérir une meilleure connaissance interne de leurs activités, mais l'impact stratégique de ses audits est considéré comme faible, notamment du fait d'un mauvais alignement des échéances du CNE avec celles du contrat d'établissement (Macarie-Florea, 2010). Sans doute, le fait que les résultats n'influencent pas concrètement les décisions des pouvoirs publics a aussi pesé sur ce faible impact à l'interne (Cytermann, 2013).

L'ÉVALUATION INTÉGRÉE À LA MODE AERES

L'AERES (Agence d'évaluation de la recherche et de l'enseignement supérieur), créée en 2007, se positionne comme une agence officielle d'expertise et assume d'emblée une mission d'aide à la décision pour les établissements… mais surtout pour le ministère. Elle privilégie une aide aux décideurs plutôt qu'une aide aux évalués et prône une évaluation intégrée qui concentre les activités préalablement menées par plusieurs institutions (dont le CNE). L'objectif poursuivi est clairement la réduction des dépenses publiques dans la mouvance du *new public management* (Cytermann, 2013).

Si la création de cette agence a été perçue comme introduisant une rupture forte, c'est essentiellement pour deux raisons, intimement liées, plaçant la France dans une configuration inédite au regard de ses partenaires européens (Cytermann, 2013 ; Musselin *et al.*, 2012 ; Musselin, 2013) :

- d'une part le caractère désormais concentré et systématique de l'évaluation externalisée, pratiquée selon un calendrier quinquennal, à la fois sur les établissements, sur les unités de recherche et sur les formations, par une seule agence ;
- d'autre part le fait que les résultats conditionnent en partie l'allocation de ressources par le ministère : dans le système d'information utilisé (SYMPA), figurent des critères d'activité et de performance issus directement des évaluations de l'AERES (chercheurs publiants ou produisants, notes des unités de recherche).

Cette forme d'industrialisation ●, associée aux réformes impactant l'allocation des ressources (avec la création parallèle de l'ANR, Agence nationale de la recherche qui centralise le financement sur projets des unités de recherche) exerce un pouvoir normatif fort : on passe **d'une logique de contrôle de la conformité *ex ante* à une logique de régulation *ex post* de la qualité**, obligeant plus ou moins explicitement toutes les parties à revoir leur fonctionnement. Parce que le ministère s'en sert et les présidences d'université aussi. L'évaluation de la qualité, au cœur des processus de régulation, joue alors un rôle d'aiguillon pour l'autonomie des universités (Cytermann, 2013).

Une enquête menée par l'équipe de Christine Musselin montre par exemple que les évaluations externes de l'AERES renforcent la gestion centralisée et permettent une répartition plus sélective des ressources au sein des établissements. Autrement dit, les notes attribuées aux unités de recherche et les projets ANR remportés par certaines unités sont utilisés pour fixer les budgets internes. Ces modalités d'évaluation deviennent alors largement incitatives et conduisent les acteurs à évoluer, voire à développer des tactiques pour se conformer aux normes, quoi qu'ils en pensent par ailleurs (Musselin *et al.*, 2012).

D'Iribarne Philippe (1989). *La logique de l'honneur : gestion des entreprises et traditions nationales.* Paris : Éd. du Seuil.

● Lors de la campagne d'évaluation 2012-2013, l'agence a notamment évalué 39 établissements, 600 unités de recherche, 195 mentions de licence, 105 spécialités de licence professionnelle, 280 mentions de master (1 164 spécialités), 12 mentions de master d'écoles d'ingénieurs, 74 écoles doctorales, et a sollicité l'appui de plus de 4 000 experts.

Lors de l'examen de la candidature de l'AERES au registre EQAR, les experts de l'ENQA ont émis certaines réserves relatives à la notation globale des laboratoires et des formations, susceptible de les disqualifier pour 4 à 5 ans et de simplifier à outrance les résultats devant être pris en compte dans les procédures d'allocation de moyens (Cytermann, 2013).

Les instances spécialisées dans l'habilitation des formations n'ont pas été épargnées par ce basculement en faveur d'une accréditation *a posteriori*. L'AERES étant amenée à évaluer les procédures d'habilitation de la commission des titres d'ingénieurs et de la commission des formations de gestion (plutôt inféodée aux accréditations internationales de type EQUIS ou AASCB) par exemple, ces instances se sont trouvées dans l'obligation de revoir leur positionnement et d'ajuster leur fonctionnement, non sans tension (Cytermann, 2013).

Ces effets d'imposition se sont trouvés amplifiés par l'usage littéral des normes européennes, motivé principalement par une forte préoccupation de rationalisation de moyens. Inscrite au registre européen EQAR depuis 2011, l'agence utilise effectivement les ESG de l'ENQA (2009), qu'elle interprète dans un premier temps comme une simple prescription, contraignant les établissements à se positionner sur certains critères prédéfinis. Ainsi le référentiel pour l'examen des maquettes de master organise le jugement selon des critères implicites qui institutionnalisent une différenciation par les moyens humains, matériels et financiers, disqualifiante pour les formations non « rentables » (dont les débouchés ne sont pas clairement identifiables) par exemple. Autrement dit, les experts de l'AERES sont attentifs aux liens avec la recherche locale et aux besoins économiques immédiats, à la pertinence de la formation en termes d'effectifs, de taux de réussite et de poursuite d'études ; ils examinent aussi les formes de suivi existantes (conseil de perfectionnement, évaluation par les étudiants, insertion) et de façon générale privilégient l'évaluation des résultats et négligent celle des objectifs et des activités. Rien n'est dit ou presque sur la composition des équipes enseignantes, l'organisation des stages, l'équilibre entre enseignements théoriques et professionnels, les volumes horaires, etc. (Agulhon, 2013).

À l'issue des Assises de l'enseignement supérieur qui se sont déroulées de juillet à novembre 2012 partout en France, les rapports Berger (2012) et Le Déaut (2013) prônent une refondation complète de l'AERES, cible de critiques trop vives pour connaître de simples ajustements. Les deux préconisent la suppression de la notation (des unités de recherche et des formations) vécue comme disqualifiante ●, invitent à simplifier les procédures, trop lourdes et redondantes. Le Déaut insiste pour que la nouvelle agence opère comme une autorité indépendante garante de la qualité et la cohérence des évaluations déléguées aux établissements et unités de recherche.

Sur la base de ces critiques, le protocole de l'AERES s'est assoupli, avec des indicateurs moins normés et une tonalité générale moins prescriptive, comme en témoigne le guide réalisé pour la campagne d'évaluation 2014-2015 (2014) qui encourage plus ouvertement les établissements à prendre en charge la production d'indicateurs cohérents avec leurs activités.

Dans cette version du guide de l'AERES, les questions d'évaluation des formations relèvent essentiellement du domaine 2 (recherche et formation), champ 2 (politique de formation initiale et continue).
Référence 3 : l'établissement suit et diffuse ses résultats, et met en œuvre des dispositifs d'évaluation et de management interne de la qualité :

- des informations précises, complètes et fiables sont régulièrement publiées sur les formations.
- les résultats de l'activité de formation sont suivis régulièrement, décrits objectivement et rendus publics.
- l'évaluation des formations et de leur contenu est effective et prise en compte par les équipes pédagogiques, les entités en charge de la formation, et l'établissement dans son ensemble.

Supprimée par la loi n° 2013-660 du 22 juillet 2013, l'AERES sera néanmoins remplacée par le HCERES (Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur) dès la publication du décret d'application correspondant (article L114-3-1 du Code de la recherche). Missionnés par la ministre de l'Enseignement supérieur, Geneviève Fioraso, pour formuler des propositions concrètes quant au fonctionnement de cette nouvelle instance, Pumain et Dardel (2014) privilégient un objectif d'amélioration de la qualité de la recherche et de la formation et repositionnent la mission d'évaluation dans la politique de site. Pour les formations proprement dites, les auteurs recommandent une procédure allégée qui va dans le sens du scénario de Le Déaut (2013) sur la délégation de compétences aux établissements : il s'agit de passer « *d'une évaluation exhaustive des maquettes de chaque diplôme à une analyse plus stratégique de l'offre de formation, pratiquée aux niveaux des sites et des établissements, et s'appuyant en particulier sur l'autoévaluation et les évaluations par les étudiants* ».

En tout état de cause, quel que soit au final le périmètre d'action de cette nouvelle agence, qu'elle prône ou non un assouplissement des modalités évaluatives, d'autres mécanismes de reddition de compte existent désormais. Le passage progressif aux RCE (responsabilités et compétences élargies) permettant d'acquérir une autonomie en matière de budget et de ressources humaines, conduit à des bouleversements bien plus profonds en termes de pilotage, et les audits de l'IGAENR réalisés dans ce cadre ne sont pas des formalités (Carassus & Dupuy, 2013).

L'ÉVALUATION DES ENSEIGNEMENTS : UNE LÉGISLATION HÉSITANTE

En France, le déploiement de l'évaluation des enseignements est marqué par une valse hésitation des pouvoirs publics depuis 1992, signe d'une certaine frilosité politique face aux bouleversements induits par des mesures qui ébranlent les valeurs mêmes de la profession universitaire, voire qui requièrent la construction d'un nouveau système de valeurs (Dejean, 2002 ; Rege Colet, 2009).

Obligatoire depuis 1997, l'évaluation des enseignements s'est développée de façon très disparate. Le rapport Dejean met en évidence des écarts de pratiques très importants à plusieurs niveaux (enseignants, départements, établissements) et des initiatives, qui quelle que soit leur envergure, ne jouent aucun rôle dans les relations entre les universités et le ministère. Quelques établissements se distinguent : la démarche qualité de Bordeaux 2, pilotée par le CRAME (Centre de recherches appliquées en méthodes éducatives) et l'approche intégrée de l'OFIP (Observatoire des formations et de l'insertion professionnelle) à l'université Lille 1 par exemple (Dejean, 2002).

Si l'évaluation des formations, plus collective, est mieux acceptée, celle des enseignements se heurte à de nombreux obstacles, liés en particulier à la crainte qu'elle ne débouche sur une évaluation des enseignants. En général, l'idée qu'elle puisse aider à la résolution des problèmes posés par l'accueil des nouveaux étudiants et constituer un élément d'une politique d'amélioration de la qualité de l'enseignement supérieur et d'une revalorisation des fonctions d'enseignement n'est pas loin d'être considérée comme subversive à l'époque (Dejean, 2002).

Les enjeux sont-ils mieux perçus aujourd'hui ? Bien évidemment, le fait qu'une agence nationale telle que l'AERES distribue des bons points a nécessairement influencé les pratiques locales : en témoignent les initiatives en matière d'auto-évaluation des formations, de suivi de cohortes d'étudiants et de diplômés et d'approches par compétences présentées lors d'un colloque de l'AERES en 2012. Dans les IUT aussi, les directions se sont mobilisées pour mettre en place des outils de

gestion de leurs formations, *via* notamment l'évaluation par les étudiants (cf. colloque de l'ADIUT en 2013) : certains misent sur les protocoles proposés par l'université (IUT Lyon 1), d'autres développent leur propre démarche qualité (IUT Strasbourg), d'autres encore investissent dans des démarches inter-IUT (IUT Lyon 2). Des symposiums sont désormais organisés aussi dans des congrès tels que l'ADMEE-Europe ou l'AIPU, favorisant une prise de distance collective sur quelques dispositifs locaux ●.

Pour autant, les universités qui communiquent sur leur expérience restent peu nombreuses, et ce sont celles qui disposent depuis plusieurs années de structures dédiées ou de services d'appui à la pédagogie investis dans l'évaluation qui sont davantage portées à partager l'avancement de leurs réflexions, signe que les dispositifs ont besoin de temps pour mûrir. Celui de l'OFIP à l'université Lille 1 et celui d'ICAP à l'université Lyon 1, déployés dès les années 1990 ● font partie des plus anciens ; parmi les initiatives plus récentes, celles du SUP de l'université Paul Sabatier à Toulouse, du SUP de l'université Joseph Fourier à Grenoble ou bien encore du CIPE à l'université de Bourgogne retiennent l'attention, par l'approche écologique et réflexive qu'elles privilégient.

● Voir en particulier l'édition 2011 du congrès de l'ADMEE-Europe, tout entier consacré au thème de l'évaluation dans l'enseignement supérieur et qui a débouché sur la publication de quelques morceaux choisis dans Romainville et al. (2012) ; plusieurs symposiums ont également traité la question de l'évaluation de l'enseignement par les étudiants, lors des éditions 2014 des congrès de l'ADMEE Europe à Marrakech et de l'AIPU à Mons.

● Les activités du CRAME à Bordeaux 2, également très tôt impliqué dans l'évaluation des enseignements, sont en cours de redéfinition suite à la fusion de trois universités bordelaises en janvier 2014.

L'évaluation des enseignements au CIPE (université de Bourgogne)
Né dans le cadre d'un projet de recherche Canada/Europe (FACDEV) en 2004, le Centre d'innovation pédagogique et d'évaluation (CIPE) pilote la réalisation des enquêtes d'évaluation des formations et des enseignements par les étudiants, rendues obligatoires et systématiques par le conseil d'administration (selon un calendrier et des procédures précis). Son action s'inscrit à plusieurs niveaux :

- l'évaluation à la demande : chacun peut contacter le CIPE pour l'évaluation d'un enseignement, d'un dispositif, d'une formation, d'une modalité particulière d'enseignement ;
- l'évaluation des dispositifs de soutien des étudiants (cf. « Plan réussite en licence ») et des propres formations pédagogiques du CIPE proposées aux enseignants ;
- l'évaluation des enseignements et des formations via des enquêtes auprès des enseignants et des données administratives.

Au CIPE, l'évaluation s'inscrit dans une logique d'accompagnement des enseignants dans leur mission d'enseignement. En produisant des informations, le CIPE permet aux enseignants de conforter ou modifier leurs pratiques, dans tous les cas, de la documenter. Pour ce faire, cette logique d'accompagnement s'appuie sur des liens étroits avec l'axe « Innovation pédagogique ». Enfin, le CIPE développe une logique de valorisation de l'investissement pédagogique des enseignants-chercheurs par le biais de publications sur l'évaluation.

(Cathy Perret, université de Bourgogne, Dijon, France)

Si l'on suppose les pratiques plus nombreuses et plus matures qu'en 2002, aucune étude ne permet de vérifier cette hypothèse et on imagine que, dans certains établissements, le sujet est encore suffisamment tabou pour que l'évaluation ne soit ni plus ni moins qu'une procédure bureaucratique. Ces évaluations sont en effet d'autant plus fragiles que l'existence de services d'appui à la pédagogie est loin d'être généralisée et que la législation en matière d'évaluation des enseignements, récemment modifiée, est loin d'afficher une volonté politique affermie.

La mention de l'évaluation par les étudiants dans la version courante du guide de l'AERES (2014) confirme que ces modalités ne sont pas reconnues comme stratégiques : elles sont associées au domaine de la réussite étudiante (pas à celui de la recherche et de la formation), sous une forme qui relativise tout lien avec une visée administrative, qui ne dit rien non plus de sa visée formative et qui signale simplement l'intérêt des étudiants pour un apprentissage de la citoyenneté.

Avec l'abolition de l'arrêté du 23 avril 2002 (le plus prescriptif à ce jour en matière d'évaluation par les étudiants, mais aussi le plus ambitieux quant à son utilisation), la tendance est d'aller vers une responsabilisation plus forte des établissements dans le choix des modalités d'évaluation à mettre en œuvre, basée notamment sur le renforcement des conseils de perfectionnement. La dimension formative de l'évaluation par les étudiants reste présente (arrêté du 9 avril 1997), mais le lien avec l'habilitation des formations disparaît. Officiellement, il s'agit plutôt de stimuler le dialogue entre les équipes pédagogiques, les étudiants et les milieux socioprofessionnels en vue d'améliorer la qualité des formations (cf. arrêté du 22 janvier 2014).

Chronologie de la législation française 1992-2014 en matière d'évaluation des enseignements
Détail des textes en vigueur sur le carnet Éduveille

NB : les références en italique ci-dessous correspondent aux textes abrogés.

Arrêté du 26 mai 1992 relatif au diplôme d'études universitaires générales, licence et maîtrise (arrêté Lang - NOR: MENZ9202416A), version révisée du 10 octobre 1995, en vigueur - Titre IV Habilitation et évaluation – Art. 24
Il introduit pour la première fois la possibilité d'une évaluation des enseignements via notamment l'appréciation des étudiants, à la demande du conseil d'administration de l'établissement.

Arrêté du 9 avril 1997 relatif au diplôme d'études universitaires générales, à la licence et à la maîtrise (arrêté Bayrou - NOR: MENU9701083A), version révisée du 15 avril 1997, en vigueur - Titre IV Habilitation et évaluation – Art. 23
Il distingue l'évaluation des formations (organisation des études) réalisée à la demande du conseil d'administration de l'établissement et évaluation des enseignements dans une visée formative (qui prend en considération l'appréciation des étudiants et dont les résultats sont destinés aux seuls enseignants).

Arrêté du 23 avril 2002 relatif aux études universitaires conduisant au grade de licence (arrêté Lang - NOR: MENS0201070A), révisé le 1er août 2011, abrogé au 1er septembre 2012, remplacé par l'arrêté du 1er août 2011.
Les articles 20 et 21 introduisaient l'obligation de mettre en œuvre l'évaluation des formations et des enseignements selon des modalités variées, permettant la participation de tous les étudiants et prévoyaient que les résultats soient communiqués au ministère dans le cadre de sa démarche contractuelle avec l'établissement et au CNE dans le cadre des évaluations périodiques qu'il conduisait.

Arrêté du 1er août 2011 relatif à la licence (NOR: ESRS1119411A), révisé le 1er septembre 2012, remplace l'arrêté du 23 avril 2002, en vigueur jusqu'au 1er septembre 2014
L'arrêté est en vigueur au moment de la rédaction de ce dossier, mais il sera abrogé au 1er septembre 2014, et remplacé par l'Arrêté du 22 janvier 2014. L'article 19 préconise la mise en place de conseils de perfectionnement pour chaque formation et une évaluation des formations et des enseignements notamment au moyen d'enquêtes régulières auprès des étudiants. La perspective est d'améliorer la qualité des formations, les résultats de ces évaluations sont destinés au conseil de la composante et au CEVU.

Arrêté du 22 janvier 2014 fixant le cadre national des formations conduisant à la délivrance des diplômes nationaux de licence, de licence professionnelle et de master (NOR: ESRS1331410A), en vigueur au 1er septembre 2014 - Titre 1er Dispositions communes aux diplômes nationaux de licence, de licence professionnelle et de master – Art. 5
L'article 5 reprend globalement la trame de l'article 19 de l'Arrêté du 1er août 2011 qu'il remplacera au 1er septembre 2014. La visée décisionnelle dans le cadre d'une démarche d'auto-évaluation est amplifiée : les évaluations servent d'appui à l'évolution de l'offre des formations et la position des conseils de perfectionnement, désormais destinataires des résultats des évaluations, est renforcée.

Ces récentes évolutions de la législation sont-elles à interpréter comme une avancée, dans la mesure où la prescription est moins forte, ou comme un recul du fait que l'évaluation des enseignements n'est plus associée aux procédures d'habilitation, ne joue qu'un rôle indirect dans les évaluations externes et dans les démarches de contractualisation, et n'intervient en aucun cas dans la carrière des enseignants-chercheurs ? Les auteurs du rapport du HCEE de 2002 abonderaient sans doute pour cette seconde lecture, mais laissons le HCERES s'installer… avant de juger.

ÉVALUATION DES PROCESSUS D'ENSEIGNEMENT-APPRENTISSAGE

Cette partie s'intéresse plus précisément au troisième volet de la démarche qualité telle qu'elle est formalisée par Hénard (2010), à savoir les dynamiques d'enseignement et d'apprentissage. Il s'agit de rendre compte de travaux récents permettant d'appréhender l'évaluation au plus près des pratiques enseignantes et étudiantes, *via* les dispositifs d'EEE (évaluation des enseignements par les étudiants) et de discuter de ce qui rapproche ou au contraire oppose l'évaluation des enseignements et celle des formations d'une part et l'évaluation des enseignements et celle des enseignants d'autre part (Romainville & Coggi, 2009 ; Younès & Romainville, 2012).

30 ANS D'EEE, 30 ANS DE TENSIONS

Malgré la poursuite d'un même objectif qualité, cette évaluation des situations d'enseignement et d'apprentissage donne lieu à un nombre considérable de dispositifs, chacun d'eux se rapportant à des valeurs différentes de ce qui fonde la qualité de l'enseignement et de la manière dont il convient de la mesurer. L'évaluation par questionnaire auprès des étudiants est sans conteste la pratique la plus répandue, soit pour recueillir leur degré de satisfaction, soit pour analyser leurs perceptions relatives à l'expérience d'ensei-

gnement. Certains pays ont ainsi déployé des enquêtes nationales alors que les établissements d'enseignement supérieur mettent partout en place des enquêtes systématiques.

En Amérique du Nord, l'évaluation des enseignements est déployée de façon plus systématique et est mieux documentée. Une étude longitudinale de Seldin *et al.* (2007, 1re édition datée de 1999) confirme que l'EEE est déployée dans la quasi-totalité des établissements et que des modalités telles que l'auto-évaluation à l'aide du dossier d'enseignement (*teaching dossier* ou *portfolio*) et l'examen du matériel d'enseignement par les pairs ont connu l'une et l'autre une forte progression en 30 ans et sont présentes respectivement dans 60 % et 40 % des universités américaines à la fin des années 1990.

L'EEE est née dans les années 1960 aux États-Unis sous la pression des associations étudiantes qui souhaitaient inciter les enseignants à se préoccuper de la qualité de leurs cours en publiant des « *black books* » ; elle est donc contemporaine d'une position plus active et engagée des étudiants dans la vie académique. Dans la décennie suivante, l'apparition des premiers services d'appui à la pédagogie a contribué à légitimer une approche réflexive de l'activité enseignante ; certains enseignants ont commencé à s'approprier la démarche à titre personnel, pour nourrir leur réflexion pédagogique. Dans les années 1980, ce sont les autorités des universités « ordinaires » qui se sont emparées de ces modalités d'évaluation, dans le double but de faire la preuve de leur valeur ajoutée en enseignement (face aux universités d'élite) et de disposer d'outils pour les arbitrages en matière de promotion et de renouvellement de contrat des enseignants. C'est donc avec les difficultés économiques apparues dans les années 1980 que la visée administrative d'aide à la décision pour la carrière s'est développée en Amérique du Nord et que les premières contestations se sont exprimées, ouvrant la voie à un courant pléthorique de recherche. Dans les faits, les deux visées, administrative et formative, semblent aujourd'hui largement confondues, mais les EEE ne sont pas mobilisées pour l'évaluation des formations : les procédures d'accréditation qui interviennent tous les 5 à 10 ans dans les établissements américains n'examinent pas (encore) la qualité pédagogique à l'aune de ces évaluations par les étudiants et il n'existe pas d'évaluations nationales des programmes de premier cycle (Bernard, 2011 ; Leeds-Hurwitz & Hoff, 2012).

En Europe, les démarches d'évaluation intégrant les points de vue des étudiants ne sont pas nées de revendications étudiantes, ni de l'institutionnalisation de pratiques enseignantes informelles. Si les premières initiatives remontent à la fin des années 1990, alors que les travaux fondateurs de Black et Wiliam (1998) sur l'évaluation formative acquéraient une large popularité, leur entrée dans les établissements s'est effectuée par le politique et en particulier par les injonctions en faveur de la qualité de l'enseignement, corollaires des travaux d'harmonisation menés dans le cadre du Processus de Bologne : un contexte qu'on peut estimer moins favorable à leur propagation.

Mais il n'est pas rare que cette étape initiale de déploiement ne s'intègre pas dans une analyse plus large de l'adéquation entre moyens et objectifs : en donnant l'illusion d'une démarche structurée, ces questionnaires, davantage considérés comme une fin qu'un moyen, limitent la démarche qualité à une évaluation des cours par les étudiants. Un peu comme « *l'arbre qui cache la forêt* » (Rege Colet, 2009). On vise l'explicitation et d'une certaine façon on renforce les non-dits sur le nombre de questionnaires conçus par les universités et supposés être renseignés par les étudiants, sur le nombre de réponses effectivement collectées, sur la non-exploitation des résultats… Malgré une littérature scientifique abondante, avec des centaines de recherches empiriques et méta-analyses depuis la fin des années 1980, la démarche dans son ensemble reste souvent caractérisée par une forte bureaucratie et un faible ancrage dans la recherche. Au final, on observe une **contradiction fondamentale entre la richesse des réflexions et la stéréotypie des pratiques** (Romainville & Coggi, 2009).

– Detroz (2008) : article de synthèse sur les études de validité de l'EEE, en particulier celles portant sur les biais ;
– Romainville & Coggi (2009) : ouvrage de réflexion sur la mise en œuvre de l'EEE dans des universités québécoises, françaises, belges, italiennes et africaines ;
– Younès (2009) : chapitre qui propose une lecture des conceptions anglo-saxonnes de l'évaluation de l'enseignement supérieur dans ses dimensions réflexives et régulatrices ;
– Bernard (2011) : ouvrage conçu comme un guide pratique, qui mêle science et expérience pour proposer des pistes d'actions en matière d'évaluation, d'amélioration et de valorisation de l'enseignement ;
– Younès & Romainville (2012) : numéro complet de la revue *Mesure et évaluation en éducation* consacré à l'EEE, avec en particulier un article conclusif sur les perspectives raisonnées de l'EEE ;
– Romainville et al. (2013) : ouvrage composé de quelques contributions au colloque 2011 de l'ADMEE Europe, qui offre une vision systémique des problématiques évaluatives à l'université (avec un chapitre de synthèse sur les paradoxes de l'EEE).

La pénétration de l'EEE comme levier stratégique de la qualité de l'enseignement peut donc être qualifiée de « lente ». Les résistances les plus communément observées au niveau des enseignants concernent d'une part la conception libérale du métier liée à la « liberté académique » dont ils bénéficient, et d'autre part l'illégitimité des étudiants, perçus comme fondamentalement incompétents, à évaluer les prestations d'enseignement. L'une et l'autre formes de résistance, d'intensité variable selon les pays ou les configurations locales, sont assimilables à la perception d'un système d'évaluation qui sert à contrôler ou à porter un jugement de valeur sur l'activité enseignante (Dejean, 2002 ; Detroz & Blais, 2012).

Pour tenter de comprendre ce que pourrait être une utilisation raisonnée et raisonnable de l'évaluation des enseignements par les étudiants, nous nous référons très largement à plusieurs publications francophones récentes qui proposent une synthèse des recherches nord-américaines et/ou européennes (Detroz, 2008 ; Romainville & Coggi, 2009 ; Younès, 2009 ; Bernard, 2011 ; Romainville *et al.*, 2013 ; Younès & Romainville, 2012) ●. Ces travaux mettent en évidence deux types de tensions constitutives de ces pratiques évaluatives (Younès *et al.*, 2013) :

– entre une visée formative liée au développement professionnel des enseignants, une visée administrative liée aux décisions de carrière et une démarche institutionnelle qui s'inscrit dans une perspective d'amélioration des formations ;
– sur les conditions de validité de l'EEE, avec des questions relatives à la fiabilité des points de vue exprimés par les étudiants, à la validité psychométrique des questionnaires et à la validité d'usage (ses effets).

L'analyse de ces tensions, à laquelle nous nous livrons maintenant, montre que la position du curseur entre une approche générique et une approche contextualisée de la qualité, entre une démarche imposée et une démarche concertée, entre une perspective formative et une perspective stratégique connaît des variations considérables tant au niveau des politiques publiques qu'à ceux des gouvernances locales et des fonctionnements facultaires.

DES ÉTUDIANTS QUI COMPTENT

Le fait que les étudiants aient leur mot à dire sur les enseignements qu'ils reçoivent n'a en soi rien de subversif. C'est davantage le fait de les considérer comme des clients à satisfaire qui présente des risques de dérive consumériste, ou bien le fait de bureaucratiser à l'excès les passations de questionnaires sans impulser de dialogue, ni de suivi. Sans miser sur le fait que la réussite est le fruit d'une co-responsabilité et d'une co-production entre l'enseignant et les étudiants. Sans problématiser leur participation à la vie universitaire ni considérer la valeur de l'expérience citoyenne (Younès & Romainville, 2012 ; Younès *et al.*, 2013).

La *student voice* dans les enquêtes nationales

Les étudiants prennent d'ailleurs au sérieux ces questions d'évaluation. La perspective de la *student voice* qui s'est considérablement propagée dans les pays anglo-saxons depuis une quinzaine d'années s'inscrit dans des pratiques militantes généralement localisées. Bien plus qu'un ralliement métaphorique, elle traduit le souhait des étudiants de peser dans les décisions qui influent sur leur vie d'étudiant et leur expérience d'étude. En Europe, l'engagement récent de l'ESU dans le projet QUEST constitue une belle illustration de cette mouvance (ESU 2012 ; ESU 2013a et b ; Galan Palomares *et al.*, 2013). Le modèle d'évaluation proposé par le syndicat étudiant représente une contribution à la construction d'une vision commune de la qualité de l'enseignement en Europe, dans le cadre des travaux menés par l'ENQA en vue de la révision des ESG.

À propos du projet QUEST (2010-2013) piloté par l'ESU
Financé en partie par la Commission européenne, le projet QUEST, nourri par un travail d'enquêtes solide, offre une vision multidimensionnelle de la qualité, composée d'une part des contenus enseignés et des modalités pédagogiques et d'autre part des équipements et services offerts et plus généralement du climat universitaire. Les étudiants apprécient de se sentir membres à part entière de la communauté universitaire (et donc partenaires des décisions) et estiment que leurs années d'études constituent une étape essentielle dans leur développement personnel et/ou citoyen.
Pour soutenir concrètement ce positionnement comme agents de la qualité (plutôt que comme clients), l'ESU a lancé en 2009 un groupe d'étudiants-experts en assurance qualité pour stimuler les échanges entre étudiants de différents pays et fournir un accompagnement aux organisations étudiantes qui souhaitent s'engager dans cette voie.

Pour autant cette initiative européenne semble encore isolée. Dans certains pays où la *student voice* existe depuis plus longtemps, des enquêtes nationales ont été déployées pour évaluer la perception qu'ont les étudiants de leur environnement d'études. Les plus connues d'entre elles, le CEQ et la NSSE, bénéficient d'une assise solide respectivement en Australie et en Amérique du Nord et sont mobilisées par les universités dans le cadre d'un processus d'amélioration de la qualité de leurs formations.

La plus ancienne d'entre elles, le CEQ (*Course Experience Questionnaire*), mesure depuis plus 20 ans la perception qu'ont les diplômés de leur expérience d'études à l'issue de leur premier cycle dans une université australienne. Le protocole d'enquête est basé sur une recherche conduite par Ramsden et Entwistle en Grande Bretagne (1981), qui distingue cinq dimensions (et vingt-quatre items) susceptibles de favoriser des apprentissages en profondeur ● . Ces dimensions sont considérées comme stables pour évaluer un programme d'étude dans son ensemble : qualité de l'enseignement ; clarté des objectifs ; quantité de travail ; méthodes d'évaluation (des étudiants) ; développement d'habiletés génériques ●.

Les données collectées par le CEQ (et par deux autres enquêtes complémentaires) ● sont publiées dans des rapports annuels mis à la disposition des étudiants. Certaines universités, telles l'université de Sydney, se servent du CEQ depuis des années comme d'un indicateur de performance pour décider des ressources à allouer à tel ou tel département. En 2008, le CEQ a été intégré aux outils décisionnels de l'administration centrale australienne dans le cadre de la gestion de son budget de soutien à l'enseignement. Il est également devenu un outil diagnostique de premier choix pour les établissements engagés dans un processus continu d'amélioration de leurs formations focalisé sur l'ingénierie pédagogique, plutôt que sur le développement professionnel des enseignants (Barrie & Ginns, 2007).

Une autre enquête longitudinale, la NSSE (*National Survey of Student Engagement*), développée par l'université d'Indiana, s'est répandue depuis le début des années 2000, aux États-Unis mais aussi au Canada, avec pour objectif d'évaluer le degré d'engagement des étudiants dans leurs études et ce qu'ils perçoivent des efforts déployés par l'institution qui les accueille pour les aider à réussir. Elle est administrée à la fin des première et dernière années académiques du premier cycle dans les *colleges* participants (603 aux États-Unis en

● Cf. Ramsden, P. & Entwistle, N.J. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology*, vol. 51 n° 3, p. 368–383.

● L'enquête NSS (*National Student Survey*) déployée depuis 2005 par Ipsos MORI repose sur les mêmes fondements théoriques que le CEQ. Elle s'adresse aux étudiants en fin de premier cycle dans les universités du Royaume-Uni. Gibbs (2010) déplore que les principaux critères relevant effectivement du niveau d'engagement des étudiants dans leurs études soient paradoxalement (devenus) facultatifs dans ces deux questionnaires.

● Le CEQ est articulé avec deux autres enquêtes, GDS (*Graduate Destination Survey*) et PREQ (*Postgraduate Research Experience Questionnaire*), visant respectivement à rendre compte de l'orientation prise à l'issue du premier cycle (poursuite d'études ou insertion professionnelle) et à évaluer l'expérience des étudiants ayant achevé des études doctorales.

2010). Quatre dimensions structurent le questionnaire, en plus des informations relatives au profil des étudiants : participation à des activités éducatives multiples ; exigences institutionnelles et nature stimulante des cours ; perception de l'environnement d'études ; estimation de leur développement individuel, en tant qu'étudiant et que personne, depuis le début du *college*.

Cette entrée par **l'engagement des étudiants** est assez originale. Qu'il s'agisse de mesurer le niveau de défi intellectuel offert par les cours, la part d'apprentissage actif et collaboratif ou bien la quantité et qualité des interactions entre enseignant et étudiants, tous ces indicateurs, testés sur plusieurs centaines d'établissement, constituent aujourd'hui une mesure fiable de l'engagement des étudiants dans leurs études et corrélativement de leur réussite, caractérisée par des indicateurs de performance mais aussi de « gain » (ce qui a changé pendant la formation), voir Pascarella *et al.* (2010). La NSSE est ainsi utilisée par les établissements pour identifier leurs faiblesses en termes d'enseignement et pour démontrer l'impact positif lié à l'introduction de certaines méthodes pédagogiques ●.

En France, on pense naturellement à l'enquête triennale de l'OVE (Observatoire de la vie étudiante) sur les conditions de vie des étudiants. Basée sur le prérequis d'une liaison entre cognition et socialisation (et donc d'une relation entre les perceptions étudiantes et le contexte d'apprentissage), elle comporte des éléments susceptibles d'appréhender la qualité de l'enseignement. Mais comme l'enquête Génération du CEREQ, elle n'a pas été conçue pour analyser et rendre publiques les variations au niveau des établissements.

Plus de 30 000 étudiants ont participé à l'édition 2010 et répondu à des questions permettant d'apprécier leur établissement (relations avec les enseignants et le personnel administratif, équipements pédagogiques, culturels, sportifs, etc.), leur formation (information sur la formation, entraide entre étudiants, intérêt des cours, disponibilité des enseignants, qualité pédagogique des enseignants) et leurs pratiques d'études (temps consacré au cours et au travail personnel, absentéisme, fréquentation de la bibliothèque, motivations, etc.). Si la granularité « établissement » n'est pas exploitée, les résultats mettent essentiellement en évidence le poids de la filière d'études : c'est l'**emprise du contexte disciplinaire**, parmi les variables externes, qui prédit le mieux les variations de perception de l'environnement d'études et de conduites adoptées par les étudiants. Des résultats qui plaident, dès lors, pour des instruments de mesure de la qualité définis au plus près des pratiques d'enseignement et d'apprentissage (Paivandi, 2012).

La fiabilité du jugement des étudiants

Certaines réticences à prendre part à des EEE peuvent être observées chez les étudiants, surtout dans des formations à effectifs réduits, quand enseignants et étudiants sont placés dans une relative proximité ou quand les enseignants ont le pouvoir d'influer significativement sur leurs parcours. Dans ces configurations particulières, la question cruciale de la préservation de l'anonymat peut alors conduire l'administration à opter pour un questionnaire papier (Romainville & Coggi, 2009). Mais d'une façon générale, les étudiants apprécient plutôt le fait qu'on leur demande leur avis et, quand ils remplissent les questionnaires, le font avec sérieux ; les questions ouvertes ne servent pas de défouloir et leur contenu est cohérent par rapport aux questions fermées (Marsh, 2007 ; Bernard, 2011).

La question de la compétence des étudiants à évaluer leurs enseignements, soulevée de façon récurrente par les détracteurs et résistants à l'EEE, paraît cependant mal posée : il ne s'agit pas de présupposer l'existence d'une compétence générique, mais de créer les conditions favorables pour un engagement des étudiants dans leurs études, l'EEE pouvant dès lors servir d'aiguillon (Detroz & Blais, 2012).

● Du fait de sa notoriété, elle a été déclinée en deux enquêtes complémentaires, la FSSE (*Faculty Survey of Student Engagement*) et la BCSSE (*Beginning Colleges Survey of Student Engagement*), avec pour objectif de mesurer respectivement les perceptions qu'ont les enseignants de l'engagement de leurs étudiants et les représentations qu'ont les néo-étudiants des attentes de l'enseignement supérieur à leur égard. Toutes ces enquêtes mettent en évidence les écarts de perception entre les sondés sur le travail à fournir en dehors des cours.

S'il est avéré que les étudiants sont influencés dans leurs jugements par leur propre rapport aux études (Michaut, 2009), ils sont également considérés comme des observateurs de première ligne, capables de fournir une rétroaction concrète dans le cadre d'une évaluation formative et dont les opinions répétées constituent des indications fiables des habiletés des enseignants pour l'administration (Centra, 1993 ; Seldin *et al.*, 2007 ; Marsh, 2007 ; Detroz, 2008 ; Romainville & Coggi, 2009 ; Bernard, 2011).

Plusieurs méthodes ont été employées pour prouver la fiabilité du jugement étudiant : certaines recherches ont montré les convergences entre les appréciations des étudiants et celles des enseignants pour un même cours ; d'autres ont souligné la cohérence des résultats collectés selon des méthodes différentes ; d'autres enfin ont établi une corrélation entre l'évaluation positive d'un enseignement et la réussite à un examen final standardisé (Detroz, 2008 ; Romainville & Coggi, 2009).

Si toutes ne présentent pas la même robustesse et si les résultats ne sont pas univoques, il existe un faisceau d'arguments convergents en faveur de la validité d'une démarche d'EEE. Parmi ces travaux, les études de biais apparaissent d'un grand intérêt pour comprendre quels sont les facteurs exogènes qui influencent positivement ou négativement l'évaluation, quelles sont les caractéristiques de l'enseignant, de l'étudiant ou de l'environnement qui interfèrent en définitive sur les réponses (Detroz, 2008) ● .

La littérature relève par exemple que le fait d'être un enseignant novice qui s'adresse à un grand groupe d'étudiants eux-mêmes novices (en première année) dans un cours obligatoire augmente la probabilité d'être moins bien évalué, comparativement à un professeur titulaire expérimenté qui enseigne un cours optionnel (choisi) à un groupe restreint d'étudiant en fin d'études.

L'impact d'autres critères peut en revanche être considéré comme marginal, même si les mythes subsistent, en particulier dans les représentations des enseignants eux-mêmes. Le *genre* par exemple intervient peu, les étudiants ne varient pas significativement dans leur jugement selon que le cours est donné par un enseignant et une enseignante. De même le fait d'être un *excellent communicant* ou de faire preuve d'*enthousiasme* pour la matière n'est pas systématiquement corrélé à une meilleure évaluation ; le même scepticisme peut être appliqué au fait d'être *sympathique* et *chaleureux*. D'une façon générale, les recherches montrent que les étudiants savent faire la différence entre une attitude et une compétence.

S'agissant des biais relatifs à la relation de travail entre l'enseignant et l'étudiant, il est également établi qu'un enseignant *indulgent dans la notation* ou *peu exigeant envers les étudiants* n'est pas mieux évalué. Les résultats sont très clairs concernant la quantité de travail demandé aux étudiants : ce sont les enseignants les plus exigeants qui obtiennent les meilleurs scores ; ils sont plus ambivalents concernant le rapport à la note espérée ou obtenue, mais globalement une surévaluation des étudiants ne garantit pas une surévaluation du cours.

Si la variance liée à la personnalité de l'étudiant n'est pas mise en évidence, sa motivation initiale pour le cours compte : ainsi il est établi que l'**intérêt pour la matière** influence à la fois les apprentissages et les résultats de l'EEE, dans une proportion qui peut être importante. En revanche, leur jugement s'avère relativement stable dans le temps : les anciens étudiants diplômés, une fois insérés dans la vie professionnelle, ne changent pas d'opinion.

De même, les contextes d'enseignement semblent impacter à la marge les évaluations : la taille de l'auditoire peut être neutralisée, et la variance disciplinaire, *a priori* légèrement défavorable aux mathématiques et aux matières scientifiques, ne permet pas de conclure

● Le jeu-questionnaire de l'université de Montréal sur la validité de l'évaluation de l'enseignement par les étudiants peut être testé avant de lire les paragraphes qui suivent…

que certaines disciplines sont systématiquement mieux évaluées que d'autres.

Les recherches montrent que certaines conditions inhérentes au dispositif d'EEE peuvent influencer positivement les réponses : le fait par exemple que l'enseignant soit physiquement présent dans la salle au moment de l'évaluation et lorsque les étudiants savent que leur jugement va compter pour la promotion de l'enseignant.

Cet inventaire des variables exogènes peut sembler troublant. S'agit-il véritablement de biais ? Sont-elles sans lien avec la qualité de l'enseignement ? L'examen de la littérature permet de conclure à une forte ambivalence, avec des recherches qui utilisent tour à tour ces résultats pour défendre ou au contraire discréditer l'EEE (Detroz, 2008). Une telle ambivalence plaide en définitive pour une meilleure prise en compte du contexte dans la définition de la qualité et pour une démarche négociée et concertée impliquant conjointement enseignant(s) et étudiants (Younès & Romainville, 2012 ; Younès *et al.*, 2013).

VERS UNE MISE EN ŒUVRE MOINS NORMATIVE

Le « pourquoi » conditionne le « comment »

L'évaluation des enseignements peut répondre à différents objectifs : fournir un diagnostic à l'enseignant, améliorer la formation des étudiants, informer les décisions administratives relatives aux enseignants, valoriser l'enseignement, éclairer les choix de cours des étudiants ou bien encore fournir des données aux chercheurs (Bernard, 2011). Mais les politiques d'évaluation ne sont pas toujours explicites quant aux objectifs poursuivis, et les frontières entre évaluation de formations, évaluation des enseignements et évaluation des enseignants d'une part et entre évaluation administrative (contrôle) et évaluation formative (accompagnement) d'autre part s'avèrent souvent poreuses (Romainville & Coggi, 2009 ; Younès & Romainville, 2012 ; Younès *et al.*, 2013).

Dans ces débats, deux positions sont observées : certains considèrent qu'une démarche de type institutionnelle ou administrative (sommative) n'est pas compatible avec une approche formative et que les visées doivent être clairement distinguées et incarnées dans des protocoles séparés. Pour d'autres, il n'existe pas de ligne de démarcation fixe, toute visée formative n'étant pas exempte de visée sommative, en particulier quand il s'agit d'améliorer les enseignements ; à l'inverse une évaluation sommative peut conduire à des ajustements au niveau des pratiques individuelles d'enseignement (Centra, 1993 ; Younès, 2009). Pour autant, insister sur ce qui sépare a peu de chance de fédérer, il est préférable de mettre l'accent sur l'adhésion des acteurs, la pertinence des critères et la qualité du suivi pour encourager une dynamique positive (Romainville & Coggi, 2009 ; Younès *et al.*, 2013).

Miser sur les convergences ne veut pas dire privilégier un seul protocole. Il semble établi qu'une évaluation formative à mi-trimestre, menée à l'initiative de l'enseignant sur l'un ou l'autre de ses cours ou sur une dimension donnée de son activité (utilisation des supports par exemple), à l'aide d'un instrument simple, oral ou écrit, a toutes les chances de l'aider à améliorer son enseignement. Il peut s'appuyer sur l'aide des conseillers pédagogiques qui élaborent souvent leurs propres grilles d'analyse ou bien utiliser des ***classroom assessment techniques*** (CATs) (Angelo & Cross, 1993) pour bénéficier d'une rétroaction indicative rapide facilitant les ajustements. La démarche dans ce cas est résolument qualitative et oblige les tiers impliqués (conseillers pédagogiques) à une certaine confidentialité (Bernard, 2011).

Les évaluations systématiques des enseignements et des formations, quelle que soit leur fréquence, sont en revanche plutôt quantitatives et interviennent à la fin des cours ou des cycles

d'études. Elles donnent lieu à une diversité de configurations. Elles peuvent par exemple être confondues : un questionnaire contrôlé par l'administration centrale, permettant de collecter des données consolidées à l'échelle de l'institution dans le cadre d'une démarche d'auto-évaluation, est complété par des questions proposées directement par les composantes qui en font alors une exploitation « locale ».

Elles peuvent aussi être séparées : l'évaluation des formations relève alors exclusivement de la responsabilité des autorités centrales ; son but n'est pas d'améliorer l'enseignement, mais de disposer d'indicateurs tangibles pour faire évoluer son offre globale de formations et négocier la reconduite des habilitations ; la granularité recherchée n'est donc pas celle des enseignements et les données agrégées sont alors de nature très diverse. Une évaluation distincte des enseignements (et des enseignants) relève alors de la responsabilité de la composante et peut s'incarner dans des protocoles plus ou moins complexes et plus ou moins normalisés au niveau de l'université, selon que l'objectif est de détecter des problèmes éventuels (« ***détecteur de fumée*** ») ou bien d'analyser le bien-fondé de tel ou tel choix pédagogique. Cette évaluation des enseignements peut aussi être réalisée à l'initiative des enseignants, afin de leur fournir une rétroaction sur l'ensemble de leurs prestations (cours magistraux, travaux pratiques, supervision des stages, encadrement des mémoires, etc.) dans la perspective d'une demande de promotion. L'enseignant peut alors être invité à choisir ses questions dans une banque d'énoncés ou sélectionner les questionnaires correspondant aux situations d'enseignement à évaluer, si ces outils sont mis à disposition par l'établissement.

Une gestion équilibrée de l'effort à l'UJF

La charte *Qualité des formations et des enseignements*, déployée à l'université Joseph-Fourier depuis janvier 2012, s'appuie sur un dispositif à trois niveaux :

- l'évaluation institutionnelle des formations est centralisée et pilotée par l'institution ;
- l'évaluation institutionnelle des enseignements est un étage intermédiaire pouvant différer d'une composante à l'autre ; il est pensé comme une évaluation indicative systématique, volontairement légère, agissant comme un « détecteur de fumée » ;
- l'analyse approfondie et confidentielle d'une situation d'enseignement est réalisée à la demande d'un enseignant ou d'une équipe enseignante ; elle est portée par les conseillers pédagogiques du service universitaire de pédagogie, ceux-ci étant soumis à une clause de confidentialité.

Ce choix est original et permet une meilleure gestion de l'effort, en réservant les moyens importants à deux situations pertinentes : lorsque l'équipe enseignante en exprime le souhait, et/ou lorsque les étudiants expriment une insatisfaction.

(Julien Douady, université Joseph-Fourier Grenoble 1, Grenoble, France).

Un exemple de ces configurations variées est donné par Berthiaume *et al.* (2011) qui, à partir d'une grille composée de quatre principes (**confidentialité, responsabilité, adaptabilité et réflexivité**) ●, montrent que le positionnement des systèmes d'EEE peut être plus ou moins formatif. Ainsi l'université de Lausanne se situe clairement dans une démarche de soutien au développement pédagogique.

● Ces quatre principes, dérivés des travaux de Centra (1993), sont détaillés comme suit. Avec la confidentialité, l'enseignant pilote son EEE et est dépositaire des données ; la responsabilité implique que l'enseignant décide quel enseignement faire évaluer et quand ; la démarche d'EEE s'adapte à l'enseignant (et non l'inverse) ; et la réflexivité induit que l'EEE formative est associée à du conseil, l'enseignant est accompagné.

Dans les deux autres universités examinées (McGill au Québec et Louvain en Belgique), les visées institutionnelle, administrative et formative sont poursuivies simultanément. Dans l'une, la répartition semble équilibrée alors que dans l'autre la visée de contrôle prédomine (mais les enseignants peuvent demander l'appui du service de pédagogie universitaire à tout moment). Une configuration est-elle plus efficace qu'une autre pour améliorer de l'enseignement ? C'est toute la question des recherches conduites par les auteurs…

Multiples dimensions et qualité psychométrique

L'idée de pouvoir miser sur quelques dimensions universelles de la qualité de l'enseignement, nous l'avons vu, doit être définitivement considérée comme illusoire. À cette étape d'élaboration des questionnaires, plusieurs méthodes peuvent être employées : s'appuyer sur les théories d'apprentissage, enquêter auprès des étudiants ou confronter les points de vue des acteurs (étudiants, enseignants, administrateurs, conseillers pédagogiques) ; toutes sont assimilables à une démarche scientifique (Romainville, 2010).

Le fait qu'aucune définition acceptable par tous, comme c'est le cas pour la recherche (nombre de publications dans des revues de premier rang, taux de citations, etc.), quoi qu'on en pense par ailleurs (Gingras, 2014), ne puisse être stabilisée, constitue clairement un frein à l'adhésion des enseignants et nourrit l'argumentaire des détracteurs de l'EEE (Romainville, 2010). C'est sans doute une des raisons pour laquelle des centaines de travaux (dès le début des années 1980 avec Cohen, Feldman et Marsh) se sont penchés sur les qualités psychométriques des questionnaires, dans l'objectif de démontrer leur validité.

On distingue ainsi les promoteurs d'un **instrument multidimensionnel complet, applicable dans toutes les situations** (Marsh par exemple) et les partisans d'une **approche globale ou spécifique**, basée soit sur quelques critères génériques soit sur une sélection de critères spécifiques exploitables dans un contexte donné (Abrami et d'Apollonia par exemple).

Les travaux fondateurs de Feldman dans les années 1970 ont conduit à la construction d'une matrice d'une vingtaine d'items regroupés en trois catégories (*presentation, facilitation, regulation*) proche de celle proposée quelques années après par Marsh et dont la notoriété perdure encore aujourd'hui. Implémentée dans un instrument standardisé mondialement connu, le SEEQ (*Students' evaluation of educational quality instrument*), la matrice de Marsh repose sur neuf dimensions déclinées en 33 items (Marsh, 2007) ●.

Les 9 dimensions du SEEQ : (1) valeur de l'apprentissage/ des études ; (2) enthousiasme de l'enseignant ; (3) organisation et clarté ; (4) interaction avec le groupe ; (5) rapport individuel avec l'enseignant ; (6) étendue de la matière présentée ; (7) examens et notation ; (8) devoirs et lectures ; (9) capacité globale à enseigner.

Parmi les instruments multidimensionnels largement répandus dans les pays anglo-saxons figure aussi l'IDEA (*Instructional development and effective assessment*), initié par Hoyt aussi dans les années 1970 et aujourd'hui administré par l'association IDEA Education. Le questionnaire, dans sa version longue, comporte trois entrées (l'enseignant, les progrès réalisés et le cours) et 47 questions prédéfinies qui présentent la particularité de donner une large place aux modalités d'apprentissage (connaissances factuelles, résolution de problème, expression orale, travail en équipe, etc.).

S'ils reconnaissent que la qualité de l'enseignement peut être multidimensionnelle, ceux qui plaident pour une mesure globale ou spécifique estiment que cette multidimensionnalité est rétive à la mesure et qu'un questionnaire unique, aussi multidimensionnel soit-il, ne peut rendre compte de la diversité des types d'enseignement : conçu le plus souvent en référence à un modèle transmissif de l'enseignement, le questionnaire s'avère peu adapté pour évaluer la qualité dans des cours aux modalités plus expérientielles (apprentissage collaboratif, par projet, etc.) par exemple. D'autres arguments sont également avancés : Abrami et d'Apollonia montrent que les évaluations des étudiants, concentrées sur des dimensions génériques telles que donner un cours, faciliter les interactions et évaluer les apprentissages, sont suffisantes pour capturer une appréciation

générale et que **la complexité des instruments constitue de surcroît un frein** à l'exploitation des résultats tant pour améliorer l'offre de formation que pour appuyer les transitions professionnelles des enseignants (Abrami *et al.*, 2007).

Gibbs (2010) propose pour sa part une alternative intéressante, basée sur quatre types de critère qui réintroduisent la biographie de l'enseignant et la perception que les étudiants ont, non pas seulement de l'activité de l'enseignant, mais aussi de leur propre activité : taille du groupe classe, niveau d'effort et d'engagement des étudiants, profil des enseignants (expérience et formation) et qualité et quantité du feedback reçu par les étudiants sur leur travail. Il observe que les variables les plus pertinentes pour qualifier le niveau d'engagement sont le niveau de défi intellectuel, l'ampleur des apprentissages actifs et collaboratifs et la quantité et qualité des interactions entre enseignant et étudiants.

Pour revenir à la typologie des « 3P » (Gibbs, 2010), on voit bien que ce sont moins les dimensions relatives à l'établissement et aux résultats des apprentissages qui cristallisent les débats psychométriques, que celles qui relèvent précisément du processus d'enseignement-apprentissage et de la relation entre l'enseignant et les étudiants, et qui sont susceptibles de générer des effets de brouillage entre évaluation des enseignements et évaluation des enseignants.

Les travaux d'Abrami *et al.* (2007), dans la mise en évidence de la qualité des enseignements, attestent également d'une **frontière ténue entre évaluation des enseignements et des enseignants**. Ils démontrent que, malgré leur variabilité, les instruments sont généralement construits à partir d'une catégorisation commune étroitement liée aux qualités de l'enseignant comme *instructeur* et comme *personne*, en lien avec son rôle de *régulateur* ; une quatrième dimension générique, relativement indépendante des trois autres, concerne l'*expertise disciplinaire*.

Toute la difficulté à cette étape de construction des questionnaires consiste à trouver un niveau de standardisation qui soit capable de rendre compte des variances contextuelles tout en satisfaisant aux besoins de comparaison. Les termes du débat entre instrument multidimensionnel *versus* instrument générique ou spécifique semblent conduire à **privilégier des instruments complexes dans une perspective formative** et à opter pour des évaluations globales quand la finalité est liée à des enjeux de carrière (Abrami *et al.*, 2007).

Contextes et validité d'usage

Pour sortir des débats interminables sur la validité « interne » des questionnaires, certains chercheurs proposent de considérer la validité de l'instrument à l'aune des effets produits par l'EEE sur l'enseignement et l'apprentissage. Ce serait alors la **capacité du dispositif d'évaluation à entraîner des changements** (attendus ou inattendus) à différents niveaux du système d'enseignement qui conférerait à l'EEE sa validité d'usage. D'autres types de preuves peuvent en effet être mobilisées, largement issues du contexte : basées sur des critères externes (convergence avec les autoévaluations de l'enseignant, note de l'étudiant à l'examen, évaluation par les pairs ou les administrateurs ou par des observateurs extérieurs, commentaires écrits des étudiants, etc.) ; axées sur la fidélité ou la cohérence (plus forte quand le nombre d'évaluateurs augmente) ; axées sur la conséquence (impact sur le développement professionnel et les pratiques enseignantes). Les auteurs signalent que cette nouvelle conception de la validité, fondée sur un questionnement centré sur l'utilité réelle de l'EEE, est d'ores et déjà actée dans les *Standards for educational and psychological testing* de l'AERA, l'APA & le NCME (1999), qui font référence en matière de psychométrie (Detroz & Blais, 2012 ; Younès *et al.*, 2013) ●.

● Ces normes ont été élaborées conjointement par trois associations : American Educational Research Association (AERA), American Psychological Association (APA) et National Council on Measurement in Education (NCME) ; la première version remonte à 1966, la version courante est datée de 1999, une nouvelle version est annoncée pour 2014.

Institutionnalisation et démarche participative

La qualité du questionnaire ne fait pas tout… Ce sont les conditions d'implémentation de l'ensemble du dispositif qui sont cruciales et en particulier son **institutionnalisation**. Si l'EEE n'est pas intégrée dans la politique de l'établissement (au moyen d'une charte par exemple ●), elle court en effet le risque d'être interprétée comme largement marquée par les points de vue des personnes en charge du dossier et son potentiel mobilisateur en sera amoindri. D'autres recommandations sont émises dans la littérature. Nous transcrivons ici celles qui ressortent des travaux de Romainville (voir notamment Romainville & Coggi, 2009 ; Younès & Romainville, 2012) :

- faire en sorte que les enseignants ne doivent pas être dépossédés, mais au contraire impliqués dans toute la démarche : définition des objectifs, procédures, critères et modalités de circulation de l'information, usages et publicités des résultats ;
- prévoir un questionnaire avec une partie commune, permettant d'obtenir des résultats consolidés à l'échelle de l'établissement, et avec une partie adaptée aux enseignements concernés ;
- veiller à se préserver de tout excès conduisant à faire de l'évaluation pour l'évaluation et à ce que la fréquence d'administration et la longueur des questionnaires soient raisonnables ;
- épargner dans un premier temps les innovations pédagogiques pour éviter de stigmatiser ce qui pourrait relever de maladresses ou de simples tâtonnements ;
- privilégier une diversité de méthodes d'évaluation (quantitative/qualitative ; pendant/à la fin des cours ; outils synthétiques/analyses plus approfondies...) pour croiser les résultats ;
- prêter attention à l'information faite aux étudiants, tant sur les motifs et les procédures que sur les effets (quelles améliorations à l'issue des évaluations ?) ;
- travailler la cohérence entre l'EEE et d'autres modalités d'évaluation, dans la perspective d'une démarche qualité plus globale visant les formations, voire l'institution ;
- ne pas négliger l'accompagnement ni le suivi, en étant par exemple attentif à la manière de communiquer des résultats négatifs…

Un modèle unique et transférable n'existe pas. Les évolutions récentes de l'EEE invitent à sortir des approches normatives voire mécanistes qui ont longtemps prédominé, pour privilégier des modèles plus participatifs, inscrits dans une temporalité variable selon la préexistence d'une « culture pédagogique ». Pour Younès *et al.* (2013), il s'agit de développer une culture de l'EEE en tant que **dispositif concerté et négocié** et d'engager tous les acteurs dans un processus de réflexivité collective, y compris les étudiants, dans le cadre d'une éducation à la citoyenneté académique. L'approche se doit d'être « écologique », dans le sens où le milieu dans lequel l'enseignement se tient doit être pris en compte car il influence le sens que les acteurs accordent à l'acte d'enseigner et d'apprendre.

Pour autant, ces démarches participatives ne sont pas garanties par le simple fait de réunir des acteurs autour d'une table, et le débat peut être largement pollué par des questions de détail. Le leadership doit être ferme, basé sur une fixation préalable des étapes et du calendrier, pour acter les décisions au fur et à mesure. Les avancées, pour respecter une certaine linéarité, requièrent chez le porteur de projet une habileté à gérer des cercles concentriques, avec des allers-retours constants entre le collectif et l'intersubjectif (Detroz, 2014). De la crédibilité du processus dépend la crédibilité du jugement, c'est-à-dire la manière dont il est reçu par les enseignants (Hurteau, 2013).

La pérennité du dispositif passe par une inscription de l'EEE dans une planification pluriannuelle explicite et par de subtils jeux d'équilibre entre des démarches

● Voir par exemple la *Charte de l'évaluation formative des enseignements et des formations par les étudiant-e-s* validée par le CEVU de l'université Lyon 1.

top down contrôlées par l'administration et des démarches *bottom up* à l'initiative des enseignants. En aval, les liens entre évaluation et décision doivent être ni trop forts pour ne pas exacerber les résistances, ni trop faibles pour éviter la démobilisation des acteurs, voire la banalisation bureaucratique. Idéalement les résultats des évaluations doivent permettre une plus grande responsabilisation des enseignants et des étudiants et encourager le perfectionnement des pratiques des uns et des autres, dans un climat privilégié de transparence (Romainville, 2010).

VERS UNE EXPLOITATION PLUS STRATÉGIQUE

Les travaux sur l'impact des EEE laissent une impression mitigée, voire paradoxale, car leur efficacité formative n'est pas mise en évidence de façon univoque par la recherche. Les raisons invoquées sont généralement de plusieurs ordres : soit la méthodologie n'est pas valide, soit l'organisation institutionnelle n'est pas adaptée (faible exploitation et suivi des résultats), soit le poids des représentations et les risques psycho-sociaux sont insuffisamment pris en compte (Younès & Romainville, 2012).

Un discours bienveillant mais des perceptions ambivalentes

La qualité des instruments reste bien entendu une question cruciale ; le fait qu'ils soient « valides » suppose que l'on ait défini ce que l'on souhaite mesurer (et pourquoi) et que l'on ait vérifié qu'ils mesurent bien ce qu'ils sont censés mesurer. À ce niveau déjà, les tensions peuvent être fortes, certains établissements optant pour la mise en œuvre d'un seul instrument supposé répondre à tous les besoins, d'autres privilégiant une multitude d'instruments qui peuvent compliquer l'interprétation des résultats. Trop longs et/ou trop nombreux, les questionnaires peuvent générer une lassitude certaine, d'autant que les énoncés (équivoques, trop vagues, focalisés sur la personnalité de l'enseignant plus que sur son activité) ne permettent pas toujours de collecter des informations exploitables (Bernard, 2011).

Pourtant, la faiblesse de l'EEE réside davantage dans les mauvais usages qui en sont faits et dans les perceptions négatives qu'elle engendre. Des recherches extrêmement critiques ont été publiées en Amérique du Nord dès les années 1990 (Cashin, Centra, Ory…), dénonçant le fait que les dispositifs sont intégrés dans la réalité universitaire sans être questionnés et qu'ils sont contrôlés par l'administration sans être utilisés dans les décisions de promotion, toujours inféodées aux performances en matière de recherche (ou pire en étant mal utilisés, quand par exemple les commentaires libres des étudiants dans les évaluations sont examinés dans les commissions de promotion).

Plusieurs ont ainsi montré que l'efficacité de l'EEE à améliorer l'enseignement n'est pas garantie **si elle ne s'accompagne pas d'actions concrètes de valorisation**. Dans un sondage intercontinental réalisé par Wright & O'Neil (1995), les évaluations pratiquées en fin de semestre auprès des étudiants sont citées par les responsables des services d'appui à la pédagogie en 34e position ; ce sont « *la reconnaissance explicite de l'enseignement dans la promotion de carrière* » et « *les gestes concrets faits par les directeurs et doyens pour valoriser l'enseignement* » qui sont considérées comme les stratégies les plus efficaces pour améliorer l'enseignement. Les travaux de Seldin *et al.* (2007) au Canada, en démontrant la complémentarité entre les EEE et d'autres modalités évaluatives telles que le dossier d'enseignement, plaident également en faveur de synergies nécessaires entre évaluation et valorisation de l'enseignement.

Au Québec, l'intérêt de la démarche est positivement perçu par les étudiants, mais ils restent dubitatifs sur le fait que leur opinion compte réellement et sont nombreux à déclarer ne pas en percevoir les effets. Les administrateurs, pour

leur part, s'intéressent moins aux forces et faiblesses de chaque enseignant qu'aux priorités d'amélioration des programmes ; pour eux, l'EEE doit s'inscrire dans une démarche planifiée plus globale, mais ils déplorent le manque de ressources humaines et financières pour soutenir leur vision et plus généralement l'absence d'une culture de l'enseignement.

D'autres recherches, en sondant les perceptions des enseignants (Centra ou Bernard), mettent en évidence des décalages entre une majorité qui estime important d'être un bon enseignant et une minorité qui pense que les établissements font preuve d'un authentique volontarisme pour une approche qualité de l'enseignement (Bernard, 2011).

D'autres encore, tant aux États-Unis qu'en Europe, rapportées par Salcin *et al.* (2012), soulignent que les enseignants ne sont pas rétifs à toute évaluation, du fait qu'ils sont le plus souvent habitués à être évalués sur leur activité scientifique, et qu'ils peuvent apprécier l'intérêt de l'EEE par rapport à d'autres outils tels que le portfolio ou le cours public ; mais ils considèrent souvent que les conditions ne sont pas réunies pour que les preuves de leur activité d'enseignement obtenues par ce canal soient comptabilisées dans une perspective d'évolution de carrière et doutent par ailleurs de la légitimité de leurs collègues et supérieurs hiérarchiques à évaluer leur activité pédagogique.

Les travaux de Salcin *et al.* (2012), menés à l'université libre de Bruxelles où l'EEE a été développée après 1968, montrent que le fait que la dimension scientifique de l'activité continue à prédominer dans la progression de carrière ne discrédite pas fondamentalement l'EEE, mais ne favorise pas non plus une prise en compte réelle des feedbacks des étudiants, tant comme moyen de développement professionnel que comme outil pour valoriser l'enseignement.

En définitive, si toutes les parties peuvent tenir un discours plutôt bienveillant à l'égard de l'EEE quand il s'agit d'améliorer l'enseignement, dans les faits les administrateurs en ont une appréhension superficielle et/ou ne disposent pas des moyens nécessaires pour mener une action véritablement stratégique ; les perceptions des enseignants sont marquées par une forte ambivalence entre la nécessité d'associer l'évaluation à la valorisation d'une part et des résistances tenaces dès que la carrière est en jeu d'autre part, quand ils ne jugent pas que l'évaluation ne sert à rien (Detroz & Blais, 2012).

Conceptions de l'enseignement et estime de soi

Le poids de ces perceptions n'est pas anodin. Peu d'études traitent précisément des impacts de l'EEE sur les pratiques enseignantes, mais celles qui existent mettent en évidence une corrélation forte entre les postures de l'enseignant et les changements provoqués par les résultats d'une évaluation.

Cette posture ne s'interprète pas seulement à une échelle individuelle, elle est aussi le fruit du contexte disciplinaire de l'enseignant, plus ou moins favorable à une appréhension collective de la qualité de l'enseignement (Nasser & Fresko, 2002). Les questionnaires conçus par les équipes pédagogiques de l'université de Bourgogne, par les choix de questions opérés, traduisent une prédominance de conceptions plus ou moins magistro-centrées de l'enseignement et une vision du développement pédagogique qui néglige assez systématiquement une meilleure connaissance des étudiants (charge de travail, difficultés) et une meilleure appréhension des éléments organisationnels contextuels qui façonnent leur environ-t (Perret, 2014).

Perret & Demougeot-Lebel, 2014

Si les enseignants, individuellement, ont l'impression que les évaluations les aident à progresser, le suivi longitudinal de 195 enseignants sur 13 ans, réalisé par Marsh (2007) au travers du SEEQ, ne montre pas d'évolution sensible de la qualité de l'enseignement, en lien avec l'expérience ; autrement dit **les enseignants jugés efficaces le sont quel que soit le contexte** alors que ceux qui sont jugés inefficaces le restent toute leur carrière durant ; ainsi les profils mis en évidence (par exemple « bien organisé mais peu enthousiaste ») s'avèrent relativement stables dans le temps.

Les travaux récents de Younès et Paivandi (2014) à l'université Blaise-Pascal-Clermont-Ferrand et à l'université de Lorraine vont dans le même sens. Alors que l'EEE permet de révéler les malentendus entre enseignants et étudiants et favorise une dynamique d'intersubjectivité à différents niveaux (prise en compte des points de vue des étudiants, mise en dialogue des perspectives enseignantes et dynamique de discussion dans l'équipe pédagogique), la nature et la profondeur des changements dans la pratique d'enseignement sont **fonction de la posture plus ou moins enthousiaste, normative ou pessimiste** de l'enseignant.

La question de l'estime de soi dans le cadre des EEE est très insuffisamment traitée par la recherche, elle permettrait sans doute de mieux comprendre certaines résistances. Les résultats peuvent en effet s'avérer particulièrement désastreux, en particulier sur les jeunes enseignants, ou dans le cas de mauvaises évaluations répétées, d'autant que l'**impact émotionnel** dure plus longtemps quand il est négatif (Barras, 2014). La sensibilité féminine semble par ailleurs plus forte : les enseignantes exprimeraient plus d'émotions négatives face à une évaluation critique et seraient moins enclines à en utiliser les résultats à des fins de régulation (notamment Kogan *et al.*, 2010). D'une façon générale, les personnes ayant développé une faible estime de soi sont particulièrement affectées par les critiques, alors que les sujets à forte estime de soi, dotés de ressources autoprotectrices plus abondantes, sont moins sensibles (Younès & Romainville, 2012).

Valeur du conseil pédagogique et potentiel réflexif de l'EEE

Malgré les limites largement commentées par la recherche, les EEE sont considérées comme des mesures valides pour diagnostiquer les forces et les faiblesses d'un enseignement. Pourtant, ce diagnostic, assurément, n'est pas suffisant pour que les pratiques changent.

Pour Romainville (2010), l'atteinte d'un objectif d'amélioration des enseignements requiert un triple suivi de l'EEE : une interprétation et mise en contexte des résultats, une communication de cette analyse aux acteurs concernés et une offre de formation et/ou d'accompagnement pour remédier aux faiblesses constatées. La communication des résultats est une étape particulièrement critique : des modalités favorisant une appropriation collective, sous la forme d'un débat entre pairs ou d'un rapport d'analyse agrégeant des données de diverse nature par exemple, peuvent nourrir positivement cette culture de l'évaluation.

Pour Centra (1993), quatre conditions doivent être réunies pour que des effets puissent être repérés à l'échelle des enseignants : disposer d'information nouvelles sur leurs points forts et ses points faibles, considérer les sources d'information comme légitimes et les informations pertinentes, savoir comment utiliser ces informations pour changer et être motivé à modifier ses pratiques.

Dès lors un système d'EEE qui respecterait quelques principes tels que ceux définis par Berthiaume *et al.* (2011) (confidentialité, responsabilité, adaptabilité et réflexivité) présenterait-il une configuration plus propice à l'évolution des pratiques ? Les recherches complémentaires menées par Dumont *et al.* (2012) à l'université de Lausanne et à

la haute école d'ingénierie et de gestion du Canton de Vaud tendent à infirmer cette hypothèse. Bien que les deux établissements aient une approche de l'EEE très différente (à visée explicitement formative pour l'un et avec une dimension de contrôle beaucoup plus marquée pour l'autre), tous les deux ont vu les utilisations volontaires de l'EEE progresser de façon significative en quelques années. Les enseignants participants, même quand ils utilisent l'EEE en autonomie, disent acquérir une meilleure conscience de ce qu'ils font, disposer d'éléments pour réfléchir à l'organisation et aux objectifs de leur enseignement, et trouvent que le dialogue avec les étudiants est stimulé. Peut-on en conclure que la valeur de l'EEE réside davantage dans la **plus-value apportée par des services d'appui** à la pédagogie solidement implantés dans les deux établissements ?

La réponse, cette fois-ci, est affirmative. L'idée que l'intervention d'un conseiller pédagogique, soit en aval au moment du feedback, soit en amont et en aval pour impulser une démarche suivie, favorise un meilleur impact des EEE est largement soutenue par des recherches empiriques (Piccinin *et al.*, 1999). Plusieurs stratégies d'amélioration peuvent être mise en œuvre, comme le montre l'exemple de l'école polytechnique de Montréal (Forest, 2009). Mais peu de travaux permettent en revanche de clarifier la valeur ajoutée distinctive de ces interventions. Dans leur méta-analyse, Penny et Coe (2004) montrent qu'une rétroaction qui dépasse le simple diagnostic pour s'inscrire davantage dans une dynamique d'accompagnement, voire de remédiation, semble être plus propice à une amélioration de l'enseignement.

Le **potentiel réflexif** de l'EEE est essentiel, même s'il ne suffit pas d'y être exposé : à certaines conditions, l'EEE soutient le développement professionnel des enseignants et peut aussi encourager une démarche métacognitive chez les étudiants. Elle s'inscrit dans une démarche interprétative et critique, avec une approche quantitative qui permet d'identifier des tendances, et une approche qualitative basée sur des modalités plus narratives, qui permet une appréhension plus profonde (Younès & Romainville, 2012).

> « *Alors que l'enseignant universitaire fonctionne sur un modèle plutôt individuel et largement implicite, avec une culture de l'évaluation intuitive et personnelle, l'EEE à orientation formative renvoie au contraire à un processus concerté et collectif qui nécessite la construction d'une autre approche plus explicite et partagée, en visant à susciter une dynamique pédagogique d'explicitation et de négociation. [...] Avec ce changement de perspective, l'essentiel est d'inscrire la pratique pédagogique non dans le cadre d'un référentiel imposé mais d'inciter les enseignants à argumenter leurs choix pédagogiques et ainsi contribuer à les approfondir comme à améliorer l'ensemble de la formation, tout en considérant l'étudiant comme partie intégrante du projet de formation.* » (Younès & Romainville, 2012).

Tous les modèles de développement professionnel sont liés à des mécanismes de réflexivité centrés sur le développement d'une personne dans ses interactions avec son environnement professionnel (Paquay *et al.*, 2010) et stimulés par la formation et le conseil pédagogiques (Rege Colet & Berthiaume, 2012). Inspirés souvent des théories du cycle d'apprentissage de Kolb (1984) et du praticien réflexif de Schön (1983), ces mécanismes passent pour les enseignants par le fait de développer leurs connaissances disciplinaires et pédagogiques, d'acquérir une meilleure conscience de la diversité des étudiants et d'adhérer à une démarche collective d'explicitation. Le développement de l'expertise consiste pourtant moins en une accumulation de ressources qu'en une « *capacité intégrative d'utiliser adéquatement ces ressources* » *in situ* ; il peut relever

d'un processus de maturation au long d'une carrière et s'accompagner d'une transformation identitaire. Se pose alors la question de savoir dans quelles conditions l'évaluation contribue à activer ce développement (Paquay *et al.*, 2010).

Pour McAlpine et Weston (2002), le fait de transformer une connaissance en action requiert au départ une réflexion intentionnelle, orientée vers l'action, et cette intentionnalité est d'autant plus forte qu'elle peut s'ancrer dans l'expérience. Aux États-Unis, c'est le *Scholarship of teaching and learning* (SoTL, traduit par « expertise en matière d'enseignement et d'apprentissage ») qui, depuis les années 1990, fournit un cadre théorique et méthodologique pour penser le développement professionnel des enseignants-chercheurs. Le praticien n'est pas réflexif tout seul : ses cours sont publics et ouverts à un examen plus ou moins systématique et plus ou moins formel ● ; il est encouragé à prendre en compte les résultats de cet examen et à présenter publiquement (dans un groupe de discussion, une conférence ou une revue) ● ce qu'il a appris (Hutchings *et al.*, 2011). Par analogie, certains plaident pour le développement d'un nouveau cadre conceptuel, étroitement lié au SoTL : le *Scholarship of assessment* (traduit par « expertise en évaluation », cf. Banta & Associates, 2002).

Pour De Ketele *et al.* (2010), la condition centrale pour que l'évaluation fasse levier est la reconnaissance : il est capital que la personne soit reconnue comme sujet et que soient valorisées ces tentatives de progression et de développement. « *Dans cette perspective, l'évaluation serait* ***moins « jugement de valeur » que « valorisation »****, c'est-à-dire « mise en valeur » des potentialités et des progrès* ».

Évaluer, améliorer et valoriser pour transformer

Les tensions relatives à l'EEE traduisent en définitive l'ambiguïté inhérente, voire consubstantielle, de ces pratiques évaluatives : vouloir privilégier une démarche formative (plutôt individuelle) tout en estimant nécessaire une valorisation de l'activité d'enseignement (et donc une appréciation « externe ») en vue d'en améliorer la qualité. Dépasser cette ambiguïté implique une **mobilisation plus forte sur le collectif**, et consécutivement une emphase plus faible sur l'individuel, car l'amélioration des pratiques enseignantes individuelles ne suffit pas à l'amélioration globale des formations. Toutes les analyses plaident en définitive, ouvertement ou en creux, en faveur d'un changement culturel qui placerait l'amélioration individuelle et collective de l'enseignement au cœur des dispositifs d'évaluation et qui s'accompagnerait d'actions concrètes de valorisation de l'activité d'enseignement.

Pour Bernard (2011) ● par exemple, il s'agit de privilégier une approche plurielle dans laquelle l'EEE n'est qu'une modalité d'évaluation parmi d'autres et qui s'incarne dans un continuum « évaluation, amélioration, valorisation » de l'enseignement (EAVE). Ce processus, systématique, doit permettre d'articuler l'évaluation d'un cours en particulier (dont les résultats intéressent directement l'enseignant) et l'évaluation de tous les cours dans toutes les disciplines (dont les résultats globaux et anonymes doivent être disponibles et partagés par des commissions ou comités de programme). L'objectif de l'EAVE n'est pas de cibler les individus pour qu'ils s'améliorent : il s'agit bien d'analyser la performance d'un programme de formation pour penser les améliorations collectives. Son déploiement oblige à évaluer de manière exhaustive toutes les dimensions (pas seulement la prestation en cours) et toutes les activités d'enseignement (pas seulement les cours magistraux), et à utiliser plusieurs sources d'information et plusieurs moyens pour évaluer l'enseignement (pas exclusivement les enquêtes par questionnaire auprès des étudiants). Il doit conduire également à mettre en place une offre de formations destinée aux nouveaux enseignants avant de les évaluer. Cette vision est également celle partagée par la National Academy of Engineering aux États-Unis (King *et al.*, 2009).

Pour Rege Colet (2009), les tensions à la fois méthodologiques et idéologiques qui sous-tendent l'EEE ne sont pas à interpréter comme une simple opposition aux principes d'une mesure de la qualité, mais à la lumière des nouvelles formes organisationnelles de l'enseignement supérieur.

● Dans certaines universités américaines, se répandent par exemple les lesson studies inspirées des pratiques japonaises : un groupe de 3 à 6 enseignants travaille ensemble pour concevoir, enseigner, analyser et améliorer un cours ; l'objectif est de varier les activités pour déterminer celles qui sont les plus efficaces en termes d'apprentissage Cerbin (2011).

● La mobilisation autour du SoTL est forte aux États-Unis même si elle ne touche pas tous les enseignants. De nombreuses associations professionnelles, régionales ou disciplinaires, regroupées notamment dans ISSOTL (International Society for the Scholarship of Teaching and Learning), organisent des rencontres périodiques et possèdent leur propre revue.

● Son ouvrage, publié dans la collection des « Guides pratiques former et se former », présente de nombreux exemples de questionnaires, de rapports et de grilles d'évaluation ainsi que des suggestions d'actions concrètes pour améliorer et valoriser l'enseignement.

15 ans d'évaluation des formations par les étudiants à l'UPS
À l'université Paul-Sabatier, l'activité d'évaluation par les étudiants fait partie des missions du service universitaire de pédagogie, également en charge de la formation pédagogique des enseignants et de l'accompagnement de projets pédagogiques innovants. Ce regroupement traduit une volonté de proposer aux enseignants une entrée unique pour l'appui à l'enseignement.
Le dispositif d'évaluation s'appuie sur le recueil de l'appréciation des étudiants par le biais de questionnaires anonymes. Une Commission paritaire organisée par chaque formation discute les résultats statistiques et formule des propositions d'amélioration dans son compte-rendu. Que pouvons-nous dire après 15 ans d'expérience ?

- les évaluations sont d'autant plus productives que les enseignants sont accompagnés tout le long de la démarche (réunions, guide d'évaluation en ligne…) ;
- les commissions paritaires offrent un lieu d'échanges directs entre enseignants et étudiants qui permet, au-delà de la collecte statistique, de régler ensemble les problèmes d'organisation (calendrier, cohérence entre UE, information des étudiants…) ;
- par contre, les questions pédagogiques nécessitent un traitement plus en profondeur ; c'est cette articulation qui conditionne l'efficacité de l'évaluation et qui est la plus difficile à mettre en place dans un contexte où l'investissement dans l'enseignement n'est pas reconnu.

(Isabelle Chênerie, université Paul-Sabatier-Toulouse 3, Toulouse, France).

L'enjeu pour les établissements, est bien de développer une approche plus stratégique de la qualité, et en particulier de la qualité de l'enseignement, à l'aide des données recueillies auprès des étudiants, mais pas seulement, et en déterminant les mesures d'accompagnement et services à fournir aux enseignants pour qu'ils améliorent leur enseignement.

Dans une vision collective, positive et proactive de l'évaluation, on peut s'interroger, avec De Ketele (2013), sur l'opportunité d'évoluer vers un CV tridimensionnel qui valorise l'ensemble des fonctions de l'enseignant-chercheur : son activité scientifique, son action pédagogique et les services rendus à la communauté régionale ou internationale…

Sans doute pourrait-on également considérer autrement le rôle des étudiants : si l'apprentissage résulte d'une co-production entre enseignant(s) et étudiants, l'évaluation ne devrait-elle pas privilégier davantage la dynamique d'enseignement-apprentissage et être moins centrée sur les compétences et attitudes des enseignants ?

BIBLIOGRAPHIE

Vous retrouverez ces références et quelques autres dans notre bibliographie collaborative en ligne, qui comprend le cas échéant des accès aux articles cités (en accès libre ou en accès payant, selon les abonnements électroniques souscrits par votre institution).

- Abrami Philip C., d'Apollonia Sylvia & Rosenfeld Steven (2007). « The dimensionality of student rations of instruction: An update on what we know, do not know, and need to do ». in Perry Raymond P. & Smart John C (dir.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Dordrecht : Springer. p. 385-456.
- ACA (2006). *Perceptions of European Higher Education in Third Countries. Final Report*. Bruxelles : Academic Cooperation Association.
- AERA, APA & NCME (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association ; American Psychological Association and National Council on Measurement in Education.
- AERES (2014). *Évaluation des établissements : référentiel de l'évaluation externe. Vague A : campagne d'évaluation 2014-2015*. Paris : Agence d'évaluation de la recherche et de l'enseignement supérieur.
- Agulhon Catherine (2013). « Accréditation-habilitation dans l'enseignement supérieur : de la distinction à la sanction ». *Carrefours de l'éducation*, vol. 36 n° 2, p. 131-148.
- Angelo Thomas A. & Cross K. Patricia (1993). *Classroom assessment techniques: a handbook for college teachers*. 2e éd. San Francisco : Jossey-Bass.
- Banta Trudy W. & Associates (2002). *Building a Scholarship of Assessment*. San Francisco, CA : Jossey-Bass.
- Barras Hervé (2014). « Impact émotionnel de l'évaluation de l'enseignement par les étudiant-e-s (EEE) chez les enseignant-e-s de la Haute école pédagogique du Valais ». in *Actes du 28e Congrès de l'Association internationale de pédagogie universitaire (AIPU), Mons, 19-22 mai 2014*. Mons : Université de Mons.
- Barrie Simon & Ginns Paul (2007). « The Linking of National Teaching Performance Indicators to Improvements in Teaching and Learning in Classrooms ». *Quality in Higher Education*, vol. 13 n° 3, p. 275-286.
- Bedin Véronique (dir.) (2009). *L'évaluation à l'université évaluer ou conseiller ?* Rennes : Presses universitaires de Rennes.
- Berger Vincent (2012). *Assises de l'enseignement supérieur et de la recherche. Rapport au Président de la République*. Paris : La documentation française.
- Bernard Huguette (2011). *Comment évaluer, améliorer, valoriser l'enseignement supérieur ?* Bruxelles : De Boeck.
- Berthiaume Denis, Lanarès Jacques, Jacqmot Christine, Winer Laura & Rochat Jean-Moïse (2011). « L'évaluation des enseignements par les étudiants (EEE) ». *Recherche et formation*, vol. 67, p. 53-72.
- Black Paul & Wiliam Dylan (1998). *Inside the Black Box: Raising Standards through Classroom Assessment*. Londres : King's College.
- Bourdin Joël (2008). *Défi des classements dans l'enseignement supérieur. Rapport d'information*. Paris : Sénat.
- Carassus David & Dupuy Elodie (2013). « Enjeux et modalités de l'évaluation de la performance globale dans une organisation publique : le cas des universités françaises ». in Romainville Marc, Goasdoué Rémi & Vantourout Marc, *Évaluation et enseignement supérieur*. Bruxelles : De Boeck. p. 249-260.
- Centra John A (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. San Francisco : Jossey-Bass.
- CNE (2003). *Livre des références. Les références de l'assurance qualité dans les établissements d'enseignement supérieur*. Paris : Comité national d'évaluation.
- Commission européenne (2014). *Report on Progress in Quality Assurance in Higher Education* [réf COM(2014) 29 final]. Bruxelles : Commission européenne.
- Coulon Alain (2005). *Le métier d'étudiant : l'entrée dans la vie universitaire*. 2e éd. Paris : Économica.

- Crozier Michel (1990). *L'évaluation des performances pédagogiques des établissements universitaires. Rapport au Ministre*. Paris : La documentation française.
- Cytermann Jean-Richard (2013). « Les réformes de l'évaluation en France dans l'enseignement supérieur et la recherche : grandes tendances, premiers effets ». in Romainville Marc, Goasdoué Rémi & Vantourout Marc, *Évaluation et enseignement supérieur*. Bruxelles : De Boeck. p. 235-247.
- Le Déaut Jean-Yves (2013). *Refonder l'université, dynamiser la recherche. Mieux coopérer pour réussir. Rapport au Premier Ministre*. Paris : Premier Ministre.
- Dejean Jacques (2002). *L'évaluation de l'enseignement dans les universités françaises*. Paris : Haut conseil de l'évaluation de l'école.
- Detroz Pascal (2008). « L'évaluation des enseignements par les étudiants : état de la recherche et perspectives ». *Revue française de pédagogie*, vol. 165 n° 4, p. 117-135.
- Detroz Pascal (2014). « Vers une démarche balisant la mise en place (ou la refonte) d'un dispositif d'évaluation des enseignements ». in *Actes du 28e Congrès de l'Association internationale de pédagogie universitaire (AIPU), Mons, 19-22 mai 2014*. Mons : Université de Mons.
- Detroz Pascal & Blais Jean-Guy (2012). « Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ? ». *Mesure et évaluation en éducation*, vol. 35 n° 3, p. 5-30.
- Donnay Jean & Romainville Marc (1996). *Enseigner à l'université : un métier qui s'apprend*. Bruxelles : De Boeck.
- Dubois Pierre (dir.) (1998). *EVALUE : Évaluation et auto-évaluation des universités en Europe*. Rapport final [contrat n° SOE2 – CT 95 – 2004, projet n° 2001]. Bruxelles : Commission européenne.
- Dumont Ariane, Rochat Jean-Moïse, Berthiaume Denis & Lanarès Jacques (2012). « Les effets de l'EEE sur le développement professionnel des enseignants et l'expérience d'apprentissage des étudiants. Comparaison de deux cas suisses ». *Mesure et évaluation en éducation*, vol. 35, n° 3, p. 85-115.
- Endrizzi Laure (2011). *Savoir enseigner dans le supérieur : un enjeu d'excellence pédagogique*. Dossier d'actualité, n° 64. Lyon : École normale supérieure de Lyon.
- ENQA (2009). *Standards and Guidelines for Quality Assurance in the European Higher Education Area*. 3e éd. Helsinki : European Association for Quality Assurance in Higher Education.
- ESU (2012). *Quest for Quality for Students. Going back to basics*. Bruxelles : European Students' Union.
- ESU (2013a). *Quest for Quality for Students. Survey on Students' Perspectives. Volume II part 1*. Bruxelles : European Students' Union.
- ESU (2013b). *Quest for Quality for Students. Institutional Analysis. Volume II part 2*. Bruxelles : European Students' Union.
- Fave-Bonnet Marie-Françoise (dir.) (2006). *Évaluer formations et enseignements dans les études supérieures : Actes du colloque interuniversitaire, Angers, 27 et 28 mars 2003*. Angers : Presses de l'université d'Angers.
- Fave-Bonnet Marie-Françoise (dir.) (2010). *L'évaluation dans l'enseignement supérieur en questions*. Paris : L'harmattan.
- Fave-Bonnet Marie-Françoise & Curvale Bruno (2013). « Approches comparées en évaluation ». in Romainville Marc, Goasdoué Rémi & Vantourout Marc, *Évaluation et enseignement supérieur*. Bruxelles : De Boeck. p. 261-272.
- Felouzis Georges (2001). *La condition étudiante : Sociologie des étudiants et de l'université*. Paris : Presses universitaires de France.
- Felouzis Georges & Hanhart Siegfried (2011). *Gouverner l'éducation par les nombres ? Usages, débats et controverses*. Bruxelles : De Boeck.
- Forest Lina (2009). « Une approche intégrée d'évaluation, d'amélioration et de valorisation de l'enseignement dans une école polytechnique ». in Romainville Marc (dir.)., *L'évaluation de l'enseignement par les étudiants. Approches critiques et pratiques innovantes*. Bruxelles : De Boeck. p. 15-34.

- Galan Palomares Fernando Miguel, Todorovski Blazhe, Kazoka Asnate & Saarela Henni (2013). *Quest for Quality for Students. A Student Quality Concept. Volume III*. Bruxelles : European Students' Union.
- Gibbs Graham (2010). *Dimensions of Quality*. York : The Higher Education Academy.
- Gingras Yves (2014). *Les dérives de l'évaluation de la recherche : du bon usage de la bibliométrie*. Paris : Raisons d'agir.
- Gomel Maurice (1996). *L'évaluation de la qualité d'un enseignement supérieur. Évaluer qui ou quoi et comment ?* Paris : Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche.
- Hadji Charles (2012). *Faut-il avoir peur de l'évaluation ?* Bruxelles : De Boeck.
- Hénard Fabrice (2010). *Learning Our Lesson. Review of Quality Teaching in Higher Education*. Paris : OCDE.
- Hénard Fabrice & Roseveare Deborah (2012). *Fostering Quality Teaching in Higher Education: Policies and Practices*. Paris : OCDE.
- High Level Group on the Modernisation of Higher Education (2013). *Improving the quality of teaching and learning in Europe's higher education institutions. Report to the European Commission*. Bruxelles : Commission européenne.
- Hurteau Marthe (2013). « Aspirer à un jugement crédible dans le cadre de l'évaluation de programme ». in Romainville Marc, Goasdoué Rémi & Vantourout Marc, *Évaluation et enseignement supérieur*. Bruxelles : De Boeck. p. 145-161.
- Hutchings Pat, Taylor Huber Mary & Ciccone Anthony (2011). *The Scholarship of Teaching and Learning Reconsidered: Institutional Integration and Impact*. San Francisco, CA : Jossey-Bass.
- De Ketele Jean-Marie (2013). « Des effets positifs et pervers des classements internationaux dans l'évaluation de la recherche et des chercheurs ». in Romainville Marc, Goasdoué Rémi & Vantourout Marc, *Évaluation et enseignement supérieur*. Bruxelles : De Boeck. p. 165-188.
- De Ketele Jean-Marie, Perisset-Bagnoud Danièle, Kaddouri Mokhtar & Wittorski Richard (2010). « Une évaluation tendue entre efficacité et transformation identitaire ». in Paquay Léopold, Wouters Pascale & Van Nieuwenhoven Catherine (dir.), *L'évaluation, frein ou levier du développement professionnel ?* Bruxelles : De Boeck. p. 49-65.
- King C. Judson, Ambrose Susan A., Arreola Raoul A. & Watson Karan (2009). *Developing Metrics for Assessing Engineering Instruction: What Gets Measured is What Gets Improved*. Washington, DC : The National Academies Press.
- King Roger, Griffiths Paul & Williams Ruth (2007). « Regulatory Intermediation and Quality Assurance in Higher Education: The Case of the Auditors ». *Oxford Review of Education*, vol. 33 n° 2, p. 161-174.
- Kogan Lori R., Schoenfeld-Tacher Regina & Hellyer Peter W. (2010). Student Evaluations of Teaching: Perceptions of Faculty Based on Gender, Position, and Rank. *Teaching in Higher Education*, vol. 15 n° 6, p. 623-636.
- Kolb, David A. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, N.J : Prentice-Hall.
- Lahire Bernard (1996). *Les manières d'étudier : enquête 1994*. Paris : La documentation française.
- Lanarès Jacques (2009). « L'évaluation-conseil dans l'enseignement universitaire en Suisse ». in Bedin Véronique (dir.), *L'évaluation à l'université : évaluer ou conseiller ?* Rennes : Presses universitaires de Rennes. p. 211-232.
- Lancelot Alain (1995). *L'évaluation pédagogique des universitaires. Rapport au Ministre de l'Enseignement supérieur et de la Recherche*. Paris : La documentation française.
- Leeds-Hurwitz Wendy & Hoff Peter Sloat (2012). *Learning Matters: The Transformation of US Higher Education*. Paris : Éditions des archives contemporaines.
- Macarie-Florea Monica (2010). « Maturation et disparition d'un dispositif d'évaluation : le Comité national d'évaluation ». in Fave-Bonnet Marie-Françoise (dir.), *L'évaluation dans l'enseignement supérieur en questions*. Paris : L'harmattan. p. 31-46.

- Marsh Herbert W. 2007. « Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness ». in Perry Raymond P. & Smart John C (dir.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Dordrecht : Springer. p. 319-383.

- McAlpine Lynn & Weston Cynthia (2002). « Reflection: Issues related to improving professors' teaching and students' learning ». in Goodyear Peter & Hativa N. (dir.), *Teacher thinking, beliefs and knowledge in higher education*. Dordr

- Mérindol Jean-Yves (200 est arrivée dans les universit *d'histoire moderne et contemporaine*, vol. 55-4bis n° 5, p. 7-27.

- Michaut Christophe (2009). « L'évaluation par les étudiants de la qualité des formations universitaires : validité, limites et conseil pour le pilotage des formations et des établissements ». in Bedin Véronique (dir.), *L'évaluation à l'université : évaluer ou conseiller ?* Rennes : Presses universitaires de Rennes. p. 175-190.

- Musselin Christine (2013). How peer review empowers the academic profession and university managers: Changes in relationships between the state, universities and the professoriate. *Research Policy*, vol. 42 n° 5, p. 1165-1173.

- Musselin Christine, Barrier Julien, Boubal Camille & Soubiron Aude (2012). *Liberté, responsabilité... et centralisation des universités*. Paris : Sciences Po.

- Nasser Fadia & Fresko Barbara (2002). Faculty Views of Student Evaluation of College Teaching. *Assessment & Evaluation in Higher Education*, vol. 27 n° 2, p. 187-198.

- Paivandi Saeed (2012). « L'appréciation de l'environnement d'études et la manière d'étudier des étudiants ». *Mesure et évaluation en éducation*, vol. 35 n° 3, p. 175-199.

- Paquay Léopold, Wouters Pascale & Van Nieuwenhoven Catherine (2010). *L'évaluation, frein ou levier du développement professionnel ?* Bruxelles : De Boeck.

- Paradeise Catherine & Thoneig Jean-Claude (2011). « Réformes et ordres universitaires locaux ». in Felouzis Georges & Hanhart Siegfried, *Gouverner l'éducation par les nombres ? Usages, débats et controverses*. Bruxelles : De Boeck. p. 33-52.

- Pascarella Ernest T., Seifert Tricia A. & Blaich Charles (2010). « How effective are the NSSE benchmarks in predicting important educational outcomes? ». *Change*, vol. 42 n° 1, p. 16-22.

- Penny Angela R. & Robert Coe (2004). « Effectiveness of Consultation on Student Ratings Feedback: A Meta-Analysis ». *Review of Educational Research*, vol. 74 n° 2, p. 215-253.

- Perret Cathy (2014). « Le caractère formatif d'un dispositif d'évaluation des enseignements par les étudiants (EEE) : une analyse de l'utilité des questionnements pour une démarche réflexive ». *Mesure et évaluation en éducation* [à paraître].

- Piccinin Sergio, Cristi Carolina & McCoy Marcia (1999). « The impact of individual consultation on student ratings of teaching ». *International Journal of Academic Development*, vol. 4 n° 2, p. 75-88.

- Pumain Denise & Dardel Frédéric (2014). *L'évaluation de la recherche et de l'enseignement supérieur*. Paris : Ministère de l'Enseignement supérieur et de la Recherche.

- Ramsden Paul (1988). « Context and Strategy ». in Schmeck Ronald R. (dir.), *Learning Strategies and Learning Styles, Perspectives on Individual Differences*. Springer US. p. 159-184.

- Rauhvargers Andrejs (2011). *Global university rankings and their impact*. Bruxelles : European University Association.

- Rauhvargers Andrejs (2013). *Global university rankings and their impact. Report II*. Bruxelles : European University Association.

- Rege Colet Nicole (2009). « L'évaluation de l'enseignement au cœur des processus d'assurance qualité : l'arbre qui cache la forêt ». in Romainville Marc & Coggi Cristina (dir.), *L'évaluation de l'enseignement par les étudiants. Approches critiques et pratiques innovantes*. Bruxelles : De Boeck. p. 191-210.

- Rege Colet Nicole & Berthiaume Denis (2012). « L'évaluation des enseignants universitaires ». *Spirale*, n° 49, p. 221-235.

Perret Cathy & Demougeot-Lebel Joëlle (2014)

- Reverdy Catherine (2014). *De l'université à la vie active*. Dossier de veille de l'IFÉ, n° 91. Lyon : École normale supérieure de Lyon.
- Romainville Marc (2009). « Une expérience d'élaboration collective de critères de qualité ». in Romainville Marc & Coggi Cristina (dir.), *L'évaluation de l'enseignement par les étudiants. Approches critiques et pratiques innovantes*. Bruxelles : De Boeck. p. 145-163.
- Romainville Marc (2010). « Vers des pratiques d'évaluation instituée de l'enseignement par les étudiants ». in Fave-Bonnet Marie-Françoise (dir.), *L'évaluation dans l'enseignement supérieur en questions*. Paris : L'harmattan. p. 143-153.
- Romainville Marc & Coggi Cristina (dir.) (2009). *L'évaluation de l'enseignement par les étudiants : Approches critiques et pratiques innovantes*. Bruxelles : De Boeck.
- Romainville Marc, Goasdoué Rémi & Vantourout Marc (dir.) (2013). *Évaluation et enseignement supérieur*. Bruxelles : De Boeck.
- Salcin Alice, Wéry France, Emplit Philippe & Robert Frédéric (2012). « Les perceptions des enseignants-chercheurs de l'université libre de Bruxelles à propos de l'évaluation des enseignements ». *Mesure et évaluation en éducation*, vol. 35 n° 3, p. 61-84.
- Schön Donald A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York : Basic Books.
- Seldin Peter (dir.) (2007). *Changing practices in evaluating teaching: a practical guide to improved faculty performance and promotion/tenure decisions*. 4e éd. Bolton, MA : Jossey-Bass.
- Thélot Claude (1993). *L'évaluation du système éducatif*. Paris : Nathan.
- Varghese, N. V. (2008). *Globalization of higher education and cross-border student mobility*. Paris : Unesco.
- Vinokur Annie (2006). « La qualité de la mesure de la qualité dans l'enseignement supérieur : essai d'analyse économique ». *Education et sociétés*, vol. 18 n° 2, p. 109-124.
- van Vught Frans A. & Ziegele Frank (dir.) (2012). *Multidimensional Ranking the Design and Development of U-Multirank*. Dordrecht : Springer.
- Wright W. Alan & O'Neil Carol (1995). « Teaching Improvement Practices: International Perspectives ». in Wright, W. Alan (dir.), *Teaching Improvement Practices: Successful Strategies for Higher Education*. Bolton, MA : Anker Publishing. p. 1-57.
- Younès Nathalie (2009). « Multidimensionnalité de l'évaluation de l'enseignement universitaire dans la littérature anglophone ». in Bedin Véronique (dir.), *L'évaluation à l'université : évaluer ou conseiller ?* Rennes : Presses universitaires de Rennes. p. 191-205.
- Younès Nathalie & Paivandi Saeed (2014). « Forme et effets d'un dispositif d'évaluation de l'enseignement par les étudiants (EEE) à orientation formative ». in *Actes du 28e Congrès de l'Association internationale de pédagogie universitaire (AIPU), Mons, 19-22 mai 2014*. Mons : Université de Mons.
- Younès Nathalie, Rege Colet Nicole, Detroz Pascal & Sylvestre Emmanuel (2013). « La dynamique paradoxale de l'évaluation de l'enseignement par les étudiants ». in Romainville Marc, Goasdoué Rémi & Vantourout Marc, *Évaluation et enseignement supérieur*. Bruxelles : De Boeck. p. 109-126.
- Younès Nathalie & Romainville Marc (2012). « Les transformations actuelles de l'EEE ». *Mesure et évaluation en éducation*, vol. 35 n° 3, p. 145-173.

Notes

Pour citer ce dossier :

Endrizzi Laure (2014). *La qualité de l'enseignement : un engagement des établissements, avec les étudiants ?* Dossier de veille de l'IFÉ, n° 93, juin. Lyon : ENS de Lyon.
En ligne : http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=93&lang=fr

Retrouvez les derniers Dossiers de veille de l'IFÉ :

- Gaussel Marie (2014). *Petite enfance : de l'éducation à la scolarisation.* Dossier de veille de l'IFÉ, n° 92, avril. Lyon : ENS de Lyon.
En ligne : http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=92&lang=fr
- Reverdy Catherine (2014). *De l'université à la vie active.* Dossier de veille de l'IFÉ, n° 91, mars. Lyon : ENS de Lyon.
En ligne : http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=91&lang=fr
- Thibert Rémi (2014). *Discriminations et inégalités à l'école*. Dossier de veille de l'IFÉ, n° 90, février. Lyon : ENS de Lyon.
En ligne : http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=90&lang=fr

Abonnez-vous aux Dossiers de veille de l'IFÉ :

http://ife.ens-lyon.fr/vst/abonnement.php

© École normale supérieure de Lyon
Institut français de l'Éducation
Veille et Analyses
15 parvis René-Descartes BP 7000 – 69342 Lyon cedex 07
veille.scientifique@ens-lyon.fr
Standard : +33 (04) 26 73 11 24
ISSN 2272-0774