IV. L'évaluation pédagogique dans les universités anglo-américaines entre engagement formatif et évaluation gestionnaire : la confusion des sentiments

(Dossier SNESUP-FSU : Engagement pédagogique et évaluation des activités pédagogiques, mars 2019)

Isabelle Luciani (mars 2019)

« While it is clear that evaluation methods themselves are changing, how much these changes reflect improvement remains unresolved" (Miller et Seldin, 2014[14]).

« Standards will be raised by improving student learning rather than by better measurement of limited learning" (Gibbs et Simpson 2004[15])

Longtemps dominées par les questionnaires étudiants, les évaluations de l'enseignement dans le monde anglo-américain, devenues fondamentales comme levier de gestion managériale des personnels et comme outil de concurrence et de benchmarking entre les universités, s'appuient aujourd'hui sur des approches « multifactorielles », emboîtées, envahissantes et ambiguës, coulant dans des dispositifs formatifs séduisants des évaluations sommatives brutales. Entre de réels bons sentiments et les pires dérives du *New Public Management* (généralisation des rétributions aux mérite, rétrogradation de salaires, licenciements…), la machine à évaluer l'enseignement présente deux facettes opposées, la première aidant à justifier, et à faire accepter, la seconde.

IV. 1 Des emboîtements d'évaluation aux finalités ambiguës

À l'évaluation étudiante, base du dispositif d'évaluation des enseignements dans le monde anglo-américain, s'ajoutent aujourd'hui les *Peer Reviews of Teaching,* associant l'observation par les pairs et l'analyse du dossier pédagogique, et l'auto-évaluation, écriture réflexive sur sa pratique enseignante, matérialisée dans le *Teaching Portfolio* aux côtés d'indicateurs comparatifs de performance, de matériel pédagogique, d'un *curriculum*. Le poids relatif de ces « sources » d'évaluation diffère d'une université à l'autre. L'université d'Austin (Texas), par exemple, recommande que les observations par les pairs ne soient utilisées que comme contrepoint pour une vue d'ensemble de l'enseignement[16]. À Transylvania University, une enquête commencée en 1998 montrait la méfiance envers les pairs comme évaluateurs, alors que l'attribution des rétributions au mérite par le seul doyen ou directeur de composante ne posait aucun problème[17].

Ces brefs exemples montrent que l'apparence de consensus sur l'évaluation des enseignements s'oppose à l'hétérogénéité réelle de ces dispositifs : les universités divergent autant sur les modalités d'évaluation que sur le poids des différents acteurs et sur leur intérêt formatif ou sommatif. L'adhésion des universités anglo-américaines à ces évaluations

---

14 J. Elizabeth MILLER, Peter SELDIN, « Changing Practices in Faculty Evaluation*:* Can Better Evaluation Make *a* Difference*? » AAUP Reports and Publication.* May-June 2014, conclusions (en ligne).

15 Graham GIBBS, Claire SIMPSON (2003). « Does your assessment support your students'learning*? », Journal of Learning and Teaching in Higher Education*,1,1., http://open.ac.uk/science/fdtl/documents/lit-review.pdf Cité par Aline Germain-Rutherford (Université d'Ottawa, directrice du centre de pédagogie universitaire), « Les modalités d'évaluation des acquis », https://sup.ups-tlse.fr/colloques/docs/pres_agr_131106.pdf, 13 novembre 2006.

16 http://www.tlataskforce.uconn.edu/docs/resources/Texas_Peer_Observation_Guidelines.pdf

17 Susan H. BARR, « Learning About Post-Tenure Review From Peer Institutions », in Christine M. LICATA, Joseph C. MORREALE, (éd.), *Post-Tenure Faculty Review and Renewal : Experienced Voice,* American Association for Higher Education, Washington, DC. 2002, p. 207-226, ici p. 214.

« multifactorielles » pratiquées aujourd'hui presque partout à titre formatif ou sommatif, et louées sur les sites internet de pratiquement tous les centres de soutien à la pédagogie des universités, doit donc être nuancée.

La différenciation entre évaluation sommative et formative, proposée notamment par Scriven en 1967 et Bloom en 1971 pour évaluer les programmes et les apprentissages[18], ne recoupe pas l'ensemble des objectifs évaluatifs (évaluation pronostique, diagnostique, didactique, certificative…). Mais elle est devenu centrale dans la problématique de l'évaluation professionnelle dès lors que la prolifération de ces formes d'évaluation, portées par la vague de « responsabilisation » du *New Public Management*, a conduit à confondre les visées de formation avec les visées de gestion des personnels.

L'évaluation formative doit éclairer les acteurs du processus enseignement-apprentissage pour les faire progresser dans la durée (« L'évaluation formative est un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage, avec l'intention de modifier la situation d'apprentissage ou le rythme de cette progression, pour apporter (s'il y a lieu), des améliorations ou des correctifs appropriés[19] »). La notion d'évaluation « formatrice », plus récente, ajoute à la dimension formative celle d'une auto-évaluation régulatrice[20].

L'évaluation sommative fait un bilan de performances (« La démarche sommative consiste à sommer (au sens algébrique du terme), c'est-à-dire additionner des éléments (j'additionne des points censés représenter des niveaux de performance) ou soustraire des éléments (je soustrais des points d'un maximum en fonction des erreurs)[21] »), utilisé, dans le cadre universitaire, à des fins gestionnaires.

Peut-on en même temps évaluer pour faire progresser et évaluer pour récompenser ou sanctionner ?

IV.1.1 - Améliorer la pédagogie ou se soumettre à des objectifs sans cesse réévalués de performance : les enjeux contradictoires de l'évaluation

L'ambiguïté des objectifs de ces évaluations contribuent à ces résistances. Les évaluations étudiantes, longtemps majoritaires et prioritaires dans les dispositifs, sont apparues très tôt à l'initiative des étudiants, mais elles changent de nature dès les années 1980, quand les gouvernances imposent, dans un contexte de crise économique renforçant le *New Public Management*, une systématisation des évaluations à visée sommative fondées sur une multiplicité de « sources » (étudiants, pairs).

IV.1.1.1 - L'évaluation, pierre angulaire de la « responsabilisation » dans le New Public Management

Au cours des années 1980, les évaluations prennent de l'ampleur avec les objectifs de « responsabilisation » des universités, prônée notamment par les législateurs et les administrateurs des États[22]. Avec l'introduction des « Post-Tenure Reviews » notamment, il faut compléter les méthodes préexistantes d'évaluation du « rendement » des professeurs titulaires, déjà à l'œuvre pour les embauches, les titularisations, les salaires, les primes ou les bourses, pour

18 Michael SCRIVEN, « The methodolody of evaluation », in Ralph W. TYLER, Robert M. GAGNE et Michael SCRIVEN (éd.), *Perspectives of curriculum evaluation, Chigago*, Rand McNally, 1967, p. 39-83 ; Benjamin S. BLOOM, J. Thomas HASTINGS, et Georges F. MADAUS, *Handbook on formative and summative evaluation of student learning*, New-York, McGraw Hill, 1971.

19 Gérard SCALLON, *L'évaluation formative des apprentissages*, Tome I : La réflexion, Québec, P.U.L., 1988, p.155, cité par le *Bulletin de la documentation collégiale*, 4, 2010.

20 Georgette NUNZIATI, « Pour construire un dispositif d'*évaluation formatrice* », *Cahiers pédagogiques*, n° 280, janvier 1990, p. 47-64.

21 Jean-Marie, DE KETELE « Ne pas se tromper d'évaluation », *Revue française de linguistique appliquée*, vol. XV, n°. 1, 2010, p. 25-37.

22 James L. APPLEGATE, Lois M. NORA, « Post-Tenure Review in Kentucky : A Clash of Cultures », in Christine M. LICATA, Joseph C. MORREALE, *op. cit.*, p. 35-49.

assurer la productivité du corps professoral, réorienter ses priorités et faciliter le licenciement. Dans cette argumentation, le prédicat est que la sécurité de l'emploi – garant de la liberté académique – et l'autonomie des universités – garant de leur formation – sont les obstacles à toute responsabilisation vis-à-vis de la qualité des prestations. Dans ce paradigme, le prestige pourrait tout au plus attirer les éléments peu nombreux d'une université d'élite, mais la nécessité de rendre des comptes obligerait à l'évaluation objectivable de l'excellence[23]. Les logiques de performance, de concurrence, de classement, d'évaluation, deviennent les garants de l'assurance qualité, par un piège rhétorique qui oppose les conditions apaisées de la production et de la transmission du savoir à la possibilité de faire « mieux » – comme si le « simple » travail pour lequel nous sommes recrutés, et le « simple » salaire de ce travail ne constituaient qu'un minimum, nous laissant *de facto* toute latitude pour faire moins ; comme si la concurrence et les récompenses étaient les seules motivations pour faire plus.

Cette évolution s'est d'abord opérée dans les années 1990 à un rythme lent. Les évaluations annuelles obligatoires et la détermination de rémunérations au mérite étaient inégalement développées selon les universités et les départements, et aboutissaient rarement à des mesures drastiques contre les personnels. À partir du milieu des années 1990, en revanche, le lien entre les évaluations et les dispositifs de gestion des carrières se durcit.

L'un des derniers exemples en date de ce tournant autoritaire concerne l'Université du Tenessee : au printemps 2018, les personnels y dénoncent l'articulation de l'UT Focus Act (*Focus on College and University Success*[24]), autorisant la réduction drastique des personnels universitaires et étudiants au sein du Conseil d'admnistration, et le durcissement de la politique d'évaluation des tenures par le Président de l'Université, Joe DiPietro : « DiPietro said the proposal is not intended as a means of eliminating tenure, but rather would provide the benefit of rejuvenating under-performing faculty, rewarding under-recognized faculty, ensuring that the review process works correctly and on occasion, removing faculty who aren't meeting university standards : "In many ways, it's good stewardship of our resources"[25] ».

De fait, ces réglementations se sont banalisées, imposant des évaluations sommatives régulières dans tous les domaines de la profession[26]. Au début des années 2000, 38% des établissements imposaient déjà des sanctions à la suite d'une mauvaise évaluation. Le congédiement est mentionné expressément dans les politiques de 25 institutions en matière d'examen après la titularisation et 9 autres établissements mentionnent d'autres sanctions possibles : rétrogradation, inadmissibilité à des fonds de promotion ou de déplacement, réduction ou gel des salaires, perte d'un congé sabbatique, programme de réorientation de carrière, réaffectation des tâches, révocation d'une affectation, placement en probation, suspension. Des récompenses, telles que les augmentations de salaire au mérite, ou des congés sabbatiques supplémentaires, sont explicitement associées à l'évaluation post-tenure dans 12 établissements (California Lutheran University, Georgia Institute of Technology, Montana State University, North Carolina State University, North Park College, Northern Arizona University, Santa Clara University, University of Arizona, University of Louisville, University of Wisconsin at Superior, Valdosta State University et Valparaiso University). Ce contexte est d'autant plus pesant que 85

23 Laure ENDRIZZI, « 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, 2014, p. 2-3.

24 https://eu.knoxnews.com/story/news/politics/2018/03/26/university-tennessee-board-focus-act-senate-haslam/460456002

25 https://eu.knoxnews.com/story/news/education/2018/02/28/university-tennessee-faculty-administrative-proposal-would-essentially-eliminate-tenure/376807002

26 Voir pour ce qui suit Cathy A. TROWER, "What Is Current Policy?", in Richard P. CHAIT, *The Questions of Tenure,* Harvard University Press, Cambridge, London, 2002, p. 32-68.

établissements admettaient le licenciement pour facteurs financiers et 71 sur arrêts de programmes (« for program discontinuances »). Dans ce dernier cas, le degré de performance mesuré lors des évaluations était décisif, d'autant plus que 23% de ces établissements étudiés avaient une définition extensive de l'urgence financière (là où l'*American Association of University Professors*, AAUP, admet une urgence financière imminente menaçant la survie de l'établissement) : pour l'Université de Floride, ce sont seulement des « circonstances financières défavorables ». Nul doute que ces chiffres se sont, depuis, aggravés.

IV.1.1.2 - L'impact croissant de l'évaluation pédagogique sur les carrières

Dans cette évolution, l'enseignement a pris une part considérable : 19% des établissements ont fait de la qualité de l'enseignement une priorité absolue dans le processus de titularisation et 46% associent les promotions aux évaluations, plaçant en priorité l'évaluation de l'enseignement (75,83%). Parmi ces établissements, 12 sont des établissements de recherche, comme l'Université de l'Iowa[27].

Les enquêtes conduites par Elisabeth Miller et Peter Seldin à dix ans d'intervalle, en 2000 puis en 2010, donnent la mesure de l'évolution[28]. Les administrateurs font face à l'augmentation des coûts de fonctionnement, au manque de financement et à la menace croissante des grandes entreprises qui se sont lancées dans l'enseignement supérieur. Les évaluations sommatives deviennent donc systématiques et potentiellement fatales (« The sudden-death implications characterize contemporary evaluation practices »)[29]. Les quelque 400 doyens de facultés d'un échantillon aléatoire de collèges d'arts libéraux qui ont répondu au questionnaire montrent que le nombre de facteurs pris en compte pour ces évaluations n'a cessé d'augmenter, et qu'ils valorisent ce qui relève de la performance visible. Pour 100% des doyens, l'enseignement est un « facteur majeur » (2% de plus qu'en 2010), contre 51% pour la recherche. Ce dernier item est cependant celui qui a le plus augmenté (ils n'étaient que 41% environ en 2000). Cette double évolution traduit l'importance acquise par la performance mesurable et promouvable dans les circuits nationaux et internationaux de la concurrence et du benchmarking : satisfaction des étudiants, publications, prix... Au contraire, si 43,8% des doyens considéraient en 2000 que l'ancienneté dans le grade était un facteur majeur de leurs évaluations, ils ne sont plus que 38,1% en 2010, tandis que 23%, chiffre stable, accordent une place majeure à l'investissement dans les institutions.

Organiser cette mesure des performances implique une machinerie capable de « ratisser » un maximum d'indicateurs, tout en essayant, sous la pression des personnels notamment, de prendre en considération les contextes (biais disciplinaires, motivation des étudiants, niveaux enseignés, effectifs…). Il en ressort des dispositifs de collecte d'informations sur les performances pédagogiques de plus en plus structurés et systématiques, impliquant trois « sources » d'information[30] :

- Les évaluations par les étudiants. En 2010, elles contribuent de manière décisive à la prise de décisions en matière de gestion des personnels pour plus de 90% des doyens sondés.
- L'auto-évaluation, dont le taux d'utilisation est passé de 58,7 à 67,6%, parallèlement à l'envol des *teaching portfolios*.
- Les observations en classe, devenues le socle du *Peer Review of Teaching.* L'utilisation des visites en classes est passée de 40,3% à 60,4% entre 2000 et 2010 . À titre comparatif,

27 I*bid.*, p. 49-51.
28 Elizabeth MILLER, Peter SELDIN, art. cit.
29 *Ibid.*, p. 1.
30 Ibid.

l'examen du programme des cours n'est une source prioritaire que pour 42% des doyens. Pour autant, le corps professoral continue à recevoir l'essentiel des commentaires sur son enseignement par le retour des questionnaires étudiants et non de l'observation des pairs.

En France, l'introduction de l'évaluation de l'enseignement a commencé à l'initiative des gouvernants, en lien avec les politiques d'assurance qualité développées par la construction de l'espace européen de l'enseignement supérieur impliquant des certifications, donc des évaluations permettant de mesurer et comparer, sur la base d'objectifs définis par des contrats. L'évaluation des enseignements s'inscrit dans un processus entamé en mai 1992 avec l'arrêté Lang, révisé en 1995, qui introduit la possibilité d'une évaluation *via* l'appréciation des étudiants, à la demande du CA d'établissement. En avril 1997 l'arrêté Bayrou permet de distinguer l'évaluation sommative des formations et l'évaluation formative des enseignements, dont les appréciations étudiantes ne sont communiquées qu'aux enseignants. Dès 2002, des textes tentent de lier l'évaluation aux moyens : l'arrêté Lang du 23 avril 2002 relatif aux études universitaires conduisant au grade de licence, révisé le 1er août 2011 et abrogé le 1er septembre 2012, prévoyait la communication de toutes ces évaluations au Ministère dans le cadre de sa démarche contractuelle. Revisitant l'arrêt du 1er août 2011 relatif à la licence (révisé en septembre 2012 et abrogé en septembre 2014), l'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, renforçait la dimension décisionnelle des évaluations, mais dans le cadre restreint de l'évolution des formations et à destination notamment des conseils de perfectionnement. L'auto-évaluation tend à s'imposer avec la meilleure prise en compte de l'enseignement dans les demandes de promotion ou de congés (source : Laure ENDRIZZI, art. cit., p. 20).

IV.1.1.3 - Sauver les meubles : quand les universitaires contribuent à l'invention du système pour se préserver du pire

L'AAUP, dès 1998-1999, a clairement perçu comment la logique de ces évaluations remettait en cause le fonctionnement de l'université comme espace de production et de transmission des savoirs : « si la norme de congédiement passe de l'"incompétence" à un "rendement insatisfaisant", comme c'est le cas dans certaines propositions actuelles, les professeurs titulaires doivent alors "satisfaire" de façon récurrente les agents administratifs plutôt que les normes fondamentales de leur profession […]. Même sur les campus où l'on ne pense pas qu'il y ait un problème avec ce que l'on appelle le "bois mort" ou les membres incompétents du corps professoral, de nombreux partisans de l'évaluation post-titularisation, ainsi que ceux qui l'adoptent dans l'espoir de prévenir des attaques plus globales et flagrantes contre la tenure, voient parfois dans cet examen un moyen de réaliser des objectifs plus vastes de gestion comme la "réduction", "une restructuration" ou "une réingénierie"[31] ».

Malgré cette conscience des risques, l'AAUP accepte de réfléchir au protocole pensant proposer des principes et des critères « objectivables » censés protéger les fondements de la« tenure ». Selon le rapport de l'AAUP de 1995 *Statement on Teaching Evaluation*[32], il paraissait judicieux que la démarche repose notamment sur le dossier d'enseignement, constitué d'une description factuelle incluant potentiellement des rapports de visite, sur des mesures de la performance pédagogique et sur la prise en compte des attentes mais aussi du soutien apporté par l'établissement (cette introduction des biais contextuels sera, nous le verrons, très difficile à réaliser). À partir des pratiques des établissements, l'AAUP distingue une quinzaine de critères fondés sur l'enseignant, les étudiants ou le « curriculum ».

31 https://www.aaup.org/report/post-tenure-review-aaup-response

32 Cathy A. TROWER, art. cit., p. 50 *sqtes* (les pages analysées du rapport de 1995 sont les pp. 133-136).

Parmi les critères axés sur l'enseignant, les évaluations alors pratiquées s'appuient d'abord sur la connaissance de la discipline et la méthode d'enseignement (plus de 60%) ou l'aptitude à la présentation, à la communication et à l'organisation des connaissance (57%). L'innovation pédagogique n'intéresse prioritairement que 14% des établissements. Parmi les critères axés sur les étudiants, la motivation et la satisfaction comptent beaucoup plus que les capacités d'apprentissage : les établissements valorisent les relations avec les étudiants à l'intérieur et à l'extérieur de la classe (69%) et la capacité à motiver et stimuler la curiosité intellectuelle (51%), tandis que la capacité d'améliorer leur compétence et leur pensée critique et analytique reste secondaire (14%). Du côté du « curriculum », on trouve le leadership par le développement de programmes et de cours (66%), l'évolution du matériel didactique (42%) et les plans de cours (30%).

Sur ces bases, la plupart des universités ont créé des groupes de travail participant à l'élaboration des dispositifs. Ils se mettent au travail pour la plupart après que les législateurs aient commencé à remettre en cause la notion de titularisation, mais avant qu'aucun mandat législatif en ce sens ait été créé. Plusieurs de ces groupes de travail ont dressé le bilan de cette participation à la réforme dans l'ouvrage de Christine Licata et Joseph Morreale, *Post-Tenure Faculty Review and Renewal: Experienced Voices,* paru en 2002[33]. Des universitaires profondément inquiets y contribuent à organiser le lien entre évaluation et carrière qui remet aujourd'hui en cause les garanties de salaires et de titularisation par la soumission à des critères de performance. En 1995, par exemple, l'Arizona Board of Regents décide de développer une politique nouvelle de gestion et d'évaluation des personnels. Le groupe de travail inquiet fait le choix de participer par peur de tout perdre (« Although the loss of tenure would be disastrous for the universities, we decided the time for confrontation had not yet arrived [34]»). Même brutalité couronnée de succès à l'Université du Texas : « Post-tenure review came abruptly to the state supported universities of Texas. […] In 1996, a state senator proposed a "two strikes and you're out" revision of tenure. His proposal – "if someone receives two substandard evaluations in a row, then that person's tenure should be revoked" – challenged the institution of tenure. Texas faculty associations, joined by university administrators, campaigned to soften this blow to the tenure system. A number of key legislators worked closely with faculty leaders and administrators to craft a compromise acceptable to all parties[35]".

L'évolution est similaire au Canada. Déjà en 1988 100% des universités canadiennes poursuivaient un but administratif par l'évaluation de l'enseignement, 94% utilisaient les étudiants comme première source d'information, essentiellement via le questionnaire, et 79% utilisaient les pairs comme source d'évaluation complémentaire[36]. En 2017, la création d'un « Institut d'excellence en éducation » a réaffirmé la définition de l'efficacité pédagogique par la culture du résultat, visant à « aligne[r] la pédagogie sur des données probantes »[37].

Toutes les sources d'information sont aujourd'hui mobilisées pour apporter des preuves de sa bonne pédagogie – au détriment peut-être de sa réelle amélioration.

---

33 Christine M. Licata, Joseph C. Morreale (éd.), *op. cit.*

34 S. Vianne McLean and Thomas Callarman, «Tracking Evolving Meanings: Five Years of Post-Tenure Review in Arizona", in Christine M. Licata, Joseph C. morreale, *op. cit.*, p. 50-65, ici p. 51.

35 Debra P. Price, Dennis Longmire, Frank Fair, Laverne Warner, Paul R. Reed, William Fleming, Jo Ann M. Duffy, « Post-Tenure Review in Texas :An Evolving Response to theLegislature's Challenge » , in Christine M. Licata, Joseph C. morreale, *op. cit.*, p. 194-206, ici p. 194.

36 Huguette Bernard , MichelTrahant, « Analyse des politiques d'évaluation de l'enseignement des universités canadiennes », The Canadian Journal of Higher Education, Vol. XVIII-3, 1988 en ligne.

37 Louis LeVasseur *et al.*, « Quels enjeux pour la critique face aux nouveaux objets de la recherche en éducation ? Quelques réflexions à partir de l'exemple québécois », *Education et sociétés* 2018/1 (n° 41), p.195-213, ici p. 12.

IV.1.2 - L'évaluation par les étudiants instrumentalisée comme indice de performance

Du *Confidential Guide to Courses* des étudiants d'Harvard en 1924 – probablement le premier répertoire de cours évalués – au développement des *Black Books* des années 1960, l'évaluation par les étudiants (EEE) se développe tout d'abord, dans l'espace anglo-américain, en lien avec l'engagement des associations étudiantes dans la vie académique[38]. Mais dans la décennie 1980 les universités, notamment les universités « ordinaires », s'emparent de ces évaluations dans un double but de gestion des difficultés économiques. Elles sont un moyen de faire la preuve de leur valeur ajoutée d'enseignement face aux universités d'élite. Elles constituent des outils d'arbitrage pour les promotions et les renouvellements de contrats.

Les universitaires cherchent aujourd'hui à pondérer le poids exclusif de ces évaluations, dont on connaît les « biais » possibles, notamment selon le contenu des questionnaires[39]. L'étude de Cara Gormally, Mara Evans et Peggy Brickman rappelle que les taux de satisfaction sont systématiquement plus bas dans certaines disciplines, par exemple en sciences et en mathématiques[40]. D'autres enquêtes ont montré que les expérimentations comme l'apprentissage actif pouvaient faire baisser les notes des étudiants[41]. D'autres encore soulignent que ces évaluations peuvent conduire les enseignants à réduire les exigences des examens pour garantir la satisfaction des étudiants[42].

Ce dernier point constitue le problème le plus marquant dans des universités sélectives à forts droits d'inscription. Ce n'est pas tant la légitimité de l'évaluation étudiante qui pose problème, que, comme le rappelle Laure Endrizzi, « le fait de les considérer comme des clients à satisfaire », car on ne devrait pas faire intervenir les étudiants dans ces évaluations « sans problématiser leur participation à la vie universitaire, ni considérer la valeur de l'expérience citoyenne[43] ». Graham Gibbs, dans *Dimensions of Quality* (2010), propose par exemple d'intégrer le niveau d'effort, l'engagement, la « qualité des étudiants ».

Or, l'impact de ces évaluations, potentiellement intégrées aux *teaching portfolios,* peut être dévastateur. Certaines maquettes, par exemple, les rendent publiques pour permettre aux étudiants futurs de choisir leurs enseignements, instaurant un effet d'inertie qui peut pérenniser les difficultés passagères d'un enseignant.

IV.1.3 - Auto-évaluation et évaluation par les pairs : entre engagement personnel et injonction sommative

Le développement de l'auto-évaluation et du *Peer Review of Teaching* peut alors apparaître comme une tentative de reprendre la main pour pondérer les évaluations étudiantes et les décisions administratives. Mais plus profondément, cette valorisation de la réflexivité repose aussi sur un courant en plein essor depuis les années 1990, et qui promeut l'intégration de la réflexivité pédagogique aux finalités du travail universitaire.

38 Voir Peter SELDIN, *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*, 2nd edition. Bolton, 1997, et Huguette BERNARD , MichelTrahant, art. cit.

39 Sur cette question, voir GIBBS Graham, *Dimensions of Quality*, York, The Higher Education Academy, 2010.

40 Cara GORMALLY, Mara EVANS, Peggy BRICKMAN, « Feedback about Teaching in Higher Ed: Neglected Opportunities to Promote Change», *Life Sciences Education*, vol. 13, 187–199, Summer 2014, p. 187-199.

41 *Ibid*., p. 188.

42 *Ibid*.

43 Laure ENDRIZZI, art. cit., p. 23. Elle renvoie notamment à Nathalie YOUNÈS et Marc ROMAINVILLE. « Les transformations actuelles de l'EEE ». *Mesure et évaluation en éducation*, vol. 35 n° 3, p. 145-173.

IV.1.3.1 - Le SoTL : promotion de l'engagement pédagogique et redéfinition de l'« enseignant-chercheur »

Déjà dans la décennie 1970 apparaissent les premiers services d'appui à la pédagogie permettant l'essor d'une pratique réflexive de l'activité enseignante. Mais c'est surtout avec le courant du « Scholarship of Teaching and Learning » (SoTL), qui se développe depuis les années 1990 dans les milieux anglophones de l'éducation, que l'enseignement supérieur devient objet de recherche pour son propre protagoniste. L'évaluation par les pairs entre ici à l'université sur des bases ambiguës et contradictoires : engagement volontaire dans la réflexion pédagogique, injonction administrative à visée sommative.

La réflexivité pédagogique fait partie des différents « scolarships » identifiés par Ernest Boyer dans sa tentative de dépasser la vieille dichotomie de l'enseignement et de la recherche par la prise en compte de la « découverte », de « l'intégration » des connaissances, de « l'application » et de « l'enseignement »[44]. Pour l'AAUP, la valorisation de plusieurs domaines permet d'alléger le poids de la recherche pure dans les services d'universitaires écrasés par la diversité des tâches. C'est aussi l'argument clé qu'avancent les universités européennes, à commencer par le Royaume Uni, avec les bénéfices et les risques que cette modification du métier comporte : se trouver réifié dans une seule de ses tâches (alors que cette approche revendique de faciliter la flexibilité des carrières universitaires). Non seulement il est des domaines de recherche vers lesquels il est difficile de revenir après des années d'abandon, mais de surcroît ces profils de carrière se retrouvent corrélés à l'identité des établissements : comme l'écrit William Locke en 2004, dans une revue consacrée à la gestion des établissements supérieurs éditée par l'OCDE, « afin que cette évolution devienne une réalité pour le plus grand nombre possible d'universitaires, des changements majeurs devront intervenir dans […] l'identité des établissements […]. Ces changements seront l'œuvre de la gestion des ressources humaines […]. Pour de nombreux établissements d'enseignement supérieur, tenter de se conformer [au] modèle [unique d'excellence universitaire] ne peut que leur faire atteindre un statut de second rang[45] ». En vérité, de telles évolutions, couronnées en 2017 par le « Higher Education and Research Act », ne peuvent qu'aboutir à « la dépossession des droits d'usage sur la tenure universitaire » au profit du pouvoir des managers et du contrôle économique des universités par le développement d'une logique « entrepreneuriale », fondée sur la concurrence et les profits[46].

À la suite des travaux d'Ernest Boyer ou encore de Lee Schulman (*Knowledge and Teaching : Foundations of the New Reform,* 1987), des fondations ont travaillé en lien avec les universités, comme la *Carnegie Academy for the Scholarship of Teaching and Learning*, ou des sociétés savantes professionnelles, comme l'*Academy of Management* et l'*American Chemical Society*. C'est dans ce contexte qu'au tournant du XXIe siècle l'*American Association of Higher Education and Accreditation* (AAHEA) a mené plusieurs projets d'examen par les pairs de

44 Ernest L. BOYER, *Scholarship reconsidered : Priorities of the Professoriate,* The Carnegie Foundation for the Advancement of Teaching, 1990 (éd. Augmentée par Drew MOSER, Todd C. REAM et John M. BRAXTON en 2015).

45 William LOCKE, « Intégrer les stratégies de recherche et d'enseignement : les implications pour la gestion et la conduite des établissements au Royaume-Uni », *Politiques et gestion de l'enseignement supérieur*, 2004/3 (no 16), p. 115-137. URL : https://www.cairn.info/revue-politiques-et-gestion-de-l-enseignement-superieur-2004-3.htm-page-115.htm Le contexte britannique est très sensible à la problématique de l'intégration de la recherche et de l'enseignement, depuis qu'en 1986 les mécanismes d'évaluation de la qualité et du financement de la recherche ont été séparés de ceux de l'enseignement, le Research Assessment Exercise (RAE) épuisant toute l'énergie des universitaires.

46 Annie VINOKUR et Corine EYRAUD, « Le "*Higher Education and Research Act 2017"* : acte de décès du service public de l'enseignement supérieur en Angleterre ? », *Droit et société*, 2018/1, n° 98, p. 113-138, p. 8 *sqtes.*

l'enseignement, dont plusieurs s'inspirent de la conceptualisation par Ernest Boyer de la science de l'enseignement[47]. Il en ressort que « l'enseignant-chercheur doit […] expérimenter et documenter – quantitativement ou qualitativement – les retombées obtenues par les modifications opérées dans sa pratique d'enseignement, puis soumettre ces résultats à ses pairs, qui pourront les commenter, les évaluer et les critiquer[48] ». Il s'agit d'un engagement personnel à visée individuelle et collective, qui peut aussi se transformer en obligation de faire la preuve de son engagement.

IV.132 - Peer Review of Teaching and Peer Observation : partager de bonnes pratiques … et en payer le prix

La procédure d'évaluation par les pairs est encouragée depuis les années 1980[49], mais elle se répand vraiment avec la valorisation de la réflexivité pédagogique. L'évaluation par les pairs est vue en effet comme un outil majeur du développement collégial de cette réflexivité, sans nul doute enrichissante. Ainsi, des colloques pédagogiques et conférences publiques centrées sur leur pratique enseignante font partie des étapes de carrière des enseignants du département d'histoire de l'Université de Standford, et des cercles d'enseignement calqués sur les cercles de qualité japonais organisent à l'Université Nebraska-Lincoln des échanges planifiés sur des questions pédagogiques[50]. Toutefois, si l'évaluation par les pairs repose sur l'idée que l'enseignement est une activité savante, elle est également évaluable à des fins sommatives[51]. Cette évaluation fait désormais partie de l'arsenal évaluatif, même si l'évaluation par les étudiants reste la forme la plus utilisée. Toutes les universités affichent sur leurs sites, généralement les pages des centres de soutien à la pédagogie, leurs principes et objectifs en la matière, des exemples de formulaires et des conseils de lecture. L'observation par les pairs constitue l'un des supports de ce *Peer Review*, qui comprend aussi l'examen du matériel pédagogique, des usages de l'innovation et des technologies d'apprentissage, et d'un « échantillonnage représentatif des preuves pertinentes[52] ». Comme le définit le site de l'Université d'Austin, « les observations du comportement en classe sont destinées à passer en revue le processus d'enseignement et sa relation possible avec

47 Voir Eugene R. RICE et Mary Deane SORCINELLI, « Can the Tenure Process Be Improved ? », in Richard P. CHAIT (éd.), *The Questions of Tenure*, Harvard University Press, Cambridge, London, 2002, p. 101-124, ici p. 110 et Debra P. PRICE, et *al.*, art. cit., p. 203 sur les sources d'inspiration du Texas.

48 Mélissa THÉRAULT, « Le mouvement du Scholarship of Teaching and Learning : y-a-t-il un chercheur dans la classe ? », *Réflexion pédagogique*, vol. 24, n°4, 2011, http://aqpc.qc.ca/sites/default/files/revue/le_mouvement_du_scholarship_of_teaching_and_learning_y_a_t-il_un_chercheur_dans_la_classe_.pdf

49 COHEN, P. A., MCKEACHIE, W. J. "The Role of Colleagues in the Evaluation of College Teaching." *Improving College and University Teaching*, 1980, 28, p. 147-154.

50 Eugene R. RICE et Mary Deane Sorcinelli, art. cit., p. 108. Sur le *Peer Review of Teaching Project* de l'Université de Nebarska-Lincoln, voir https://peerreview.unl.edu

51 Le *Guide to Best Practices* de l'Université de Takoma (Evaluating_Teaching_P.T._Guide_Draft_Kalikoff_TLC_UWS, document en ligne) renvoie ici, pour expliquer le *Peer Review*, aux travaux de David GOSLING, « Collaborative Peer-Supported Review of Teaching », in J SACHS et M. PARSELL (éd.), *Peer Review of Learning and Teaching in Higher Education. International Pespectives*, Springer, 2014, p. 13-22 et Kalish, A, *Evaluation of teaching: Recommendations for developing an effective, reliable, and humane policy*, 2015. Gosling s'intéresse notamment au cas anglais (où le « Subject Review » a introduit l'observation par les pairs) et à la manière de dépasser l'opposition de l'observation formative à l'observation sommative.

52 « Formative Peer Review of Teaching » de l'University of British Columbia, https://ctlt.ubc.ca/programs/all-our-programs/peer-review-of-teaching-2/ (« a representative sampling of the relevant evidence »).

l'apprentissage. L'accent est mis sur les comportements verbaux et non verbaux de l'instructeur et des élèves en classe[53] ».

En amont de l'observation (généralement répétée, à des rythmes très différents d'une université à l'autre), les dispositifs prévoient des rencontres (*Peer-to-Peer Interviews*) au cours desquelles l'évaluateur intégrera les éléments de contexte à son observation (effectifs des classes, niveaux d'enseignement, problématiques propres à la discipline, parcours pédagogique de l'enseignant…). Les dispositifs les plus « formatifs » peuvent prévoir à ce stade de planifier des expérimentations pédagogiques au cours des séances. Les protocoles sont en général très précis, définissant l'arrivée de l'observateur, sa présentation à la classe – qu'il ne doit pas perturber –, l'endroit où il s'assoit pour observer au mieux les interactions – ce qui implique qu'un plan de salle ait été discuté lors des entretiens préliminaires –, sauf dans les cas d'observations inoppinées, prévues par exemple par l'université d'Austin[54]. Pendant l'observation, l'évaluateur peut suivre un questionnaire ou produire librement son analyse. L'observation doit être suivie d'un protocole de normalisation dont les modalités divergent, de l'informel au formel, du volontaire au disciplinaire. La rétroactivité doit déboucher *a minima* sur un « plan de perfectionnement ».

Il y a donc, en dépit d'un cadre général commun souvent trompeur sur son degré de maturité réel du fait de l'abondance des détails factuels décrits dans les dispositifs, de grandes diversités dans les mises en œuvre de cette procédure, inégalement acceptée par les personnels, d'une université ou d'un département à l'autre.

La composition des comités d'observation, par exemple, et le responsable de ces comités, peuvent relever de choix différents : collègues du même département ou de la même université, évaluateurs extérieurs choisis *ad hoc* par le directeur du département ou le Doyen, panel d'experts ? À Austin, l'observation sommative par les pairs est constituée d'un comité évaluant séparément, et comprenant en fait des représentant des étudiants et de l'administration aux côtés du corps professoral[55]. En Nouvelle-Zélande, le rapport *Peer Observation and Review of teaching and learning* de Canterbury Christ Church University, daté de juin 2013, détaille à l'envie un protocole d'observation évaluative qui met en avant un « Scholarly Dialogue Model » promouvant l'entraide et la réflexivité[56]. Le dispositif a été mis en place en 2004 pour fournir une rétroaction officielle sur l'enseignement. Le changement le plus notable depuis a été son élargissement : les évaluations peuvent intégrer, comme à Austin, des étudiants et d'autres membres du personnel de l'Université. Outre les pairs, l'observation de l'enseignement peut aussi être aujourd'hi le fait de *student ambassadors* (SALTs) ou de représentants des services de soutien à la pédagogie, à des rythmes encore indécis (participer au moins une fois par an est la règle).

Il est vrai que les objectifs sont également divers et parfois hésitants. À l'Université de Washington, la réglementation exige un examen de l'enseignement par les pairs mais laissent les objectifs et les pratiques à la discrétion des départements : évaluation formative et/ou sommative ? choix d'un pair du même département, de la même discipline, du même rang ? extension aux non titulaires[57] ? D'autres universités ont déjà fait des choix plus tranchés, comme l'Université du Texas à Austin, qui prend en compte les observations lors des promotions et des

53 http://www.tlataskforce.uconn.edu/docs/resources/Texas_Peer_Observation_Guidelines.pdf, p. 3-4.

54 Voir notamment Erica CAMPBELL, « The value of the non-evaluative: Rethinking faculty observation », in The Graduate Center Learning Collective City University of New York and Hilarie Ashton (éd.), *Structuring equality: A handbook for student centered learning and teaching practices*, 2017, p. 109-130.

55 http://www.tlataskforce.uconn.edu/docs/resources/Texas_Peer_Observation_Guidelines.pdf, p. 8.

56 https://www.canterbury.ac.uk/learning-and-teaching-enhancement/docs/rapport.pdf, 2013.

57 http://www.washington.edu/teaching/teaching-resources/assessing-and-improving-teaching/evaluation/peer-review/

titularisations depuis une note de service de mai 1995 imposant la mesure de la performance pédagogique par les étudiants et les pairs. Une promotion nécessite 2 à 4 observations, un suivi post-tenure suppose au minimum, pour le département d'histoire de l'art par exemple, un examen complet tous les six ans[58].

La tendance est à cette obligation d'observation sommative, alors même que la recherche en pédagogie est divisée sur sa confusion avec le formatif.

The University of British Columbia, à Vancouver, est un parfait exemple de cette confusion entretenue[59]. Le site du centre de soutien à la pédagogie revendique la valeur formative des *Peer Reviews* qualifiés de « coaching par les pairs », et offre du matériel pédagogique, notamment des vidéos, pour aider les enseignants. Mais la même page contient le lien vers le volet sommatif, employé pour « guider les décisions concernant le mérite, le renouvellement du mandat, la promotion et la permanence[60] ». Comment la gratuité heuristique de l'évaluation formative (tenter des expérimentations, tirer profit des « situations critiques »…) peut-elle aller de pair avec les enjeux majeurs d'une évaluation de carrière, quand bien même (c'est le cas ici) le mentor et l'examinateur sommatif sont distincts dans la procédure ? Le rapport du groupe de travail de l'université sur le *Summative Peer Review* souligne que « les équipes d'examen par les pairs devraient idéalement (« ideally ») comprendre des examinateurs qui ont reçu une formation ou possèdent une expertise pertinente[61] ». Si de tels vœux doivent être formulés « dans l'idéal », on comprend que les réserves affleurent sur cette pratique. Malgré son zèle à la mettre en place, l'université de Waterloo reconnaît ainsi que « reservations […] are understandable – reservations, for example, related to standards of evaluation, the skill of reviewers, reliability and validity of information gathered, privacy, vulnerability, and the time involved in conducting review[62] ».

IV.133 - Le Teaching Portfolio comme auto-évaluation : réflexivité pédagogique et preuves de performances

En 2005, Peter Seldin définit doublement le portfolio comme un ensemble de documents qui attestent des performances de l'enseignement, suffisamment souple pour être utilisé pour la titularisation et la promotion mais aussi pour stimuler et structurer l'auto-réflexion sur les domaines à améliorer[63]. Cet outil fait son entrée véritable sur les campus au début des années 1990[64].

À première vue, il en ressort un mélange séduisant d'engagement personnel et de volonté de partage, épaulés presque artisanalement par les guides mis en ligne par les centres de soutien à la pédagogie des universités, assortis de quelques ouvrages de références et de liens vers des modèles. L'Université du Massachusetts à Amherst met à la disposition du corps professoral un manuel intitulé *Preparing a Teaching Portfolio* (Fran Mues, Mary D. Sorcinelli, 2000), contenant des informations pratiques étape par étape, et des extraits des portfolios préparés par des professeurs du campus[65]. Cette approche illustre bien le concept clé de Lee Shulman : « A scholarship of teaching will entail a public account of some or all of the full act of teaching-

58 http://www.tlataskforce.uconn.edu/docs/resources/Texas_Peer_Observation_Guidelines.pdf
59 https://ctlt.ubc.ca/programs/all-our-programs/peer-review-of-teaching-2/.
60 https://ctlt.ubc.ca/programs/all-our-programs/ubc-peer-review-of-teaching-initiative/.
61 http://ore.educ.ubc.ca/files/2013/03/UBC-Principles-of-Summative-Peer-Review.pdf, 2013.
62 https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/professional-development/reviewing-teaching/peer-review-teaching.
63 Voir Peter SELDIN, *The Teaching Portfolio, op. cit.*
64 Voir Eugene R. RICE et Mary Deane SORCINELLI, art. cit., p. 109-110.
65 *Ibid.,* p. 109. Le manuel est en ligne : https://cndls.georgetown.edu/media/documents/teachingportfolio.pdf .

vision, ing-vision, design, enactment, outcomes, and analysis-in a manner susceptible to critical review by the teacher's professional peers and amenable to productive employment in future work by members of that same community[66] ».

En règle générale, on trouve dans le portfolio l'exposé d'une réflexion sur sa pratique et ses expérimentations, souvent qualifiée de « philosophie[67] », des revues de cours (objectifs, bilans, évaluations…), des supports (programmes, matériel didactique, devoirs…) et le *curriculum* (publications, prix, projets)…

Le *Center for Teaching* de l'université de Virginie offre un exemple intéressant de ce mélange entre un engagement personnel dont l'amélioration de la pratique est la seule fin, et une injonction institutionnelle visant à faire la preuve de son excellence dans un cadre sommatif. On y trouve des modèles respirant l'engagement sans arrière-pensée dans la pédagogie, comme cette professeure de Littérature qui explique dans ses « Reflexions » qu'elle a commencé son portfolio « principalement […] pour reconsidérer ce qu'[elle] a enseigné », et a découvert presque par hasard qu'il pouvait « mener à d'autres conversations avec des collègues et des étudiants[68] ». Le centre de soutien à la pédagogie de l'université, qui organise des ateliers, confirme que les portfolios doivent servir principalement à l'auto-réflexion et à l'amélioration personnelle[69]. Si ce n'est que la même université utilise le portfolio à des fins sommatives de gestion, par exemple pour la négociation annuelle des salaires, comme l'indique une autre page du site : « An annual performance review of all faculty members is conducted as part of the salary setting process. This review includes student evaluations […], supplemented in appropriate cases by teaching portfolios, peer attendance of classes, or other measures of teaching performance[70] ». De fait, dans un autre portfolio modèle, un enseignant en informatique, peu loquace sur sa philosophie, insère tous ses prix d'excellence (dont le « Professor of the Year award ») et tous les liens vers ses évaluations par les étudiants, questionnaires portant sur la clarté du cours, l'adéquation entre la somme de travail demandée et les crédits obtenus, la disponibilité de l'enseignant, mais aussi son degré de maîtrise disciplinaire ou encore si « As a teacher, this instructor was better than most others in this School », ainsi que des commentaires libres– « Wow, this class crushed my soul »[71]…

Dans certaines universités, la réflexivité imposée enserre véritablement le quotidien et tient presque du divan de l'analyste, tel le rapport du *Center for Teaching Support & Innovation* de l'Université de Toronto qui préconise une évaluation hebdomadaire[72], tandis que d'autres recommandent des annotations et des enregistrements audio après chaque cours…

66 SHULMAN, Lee S., « Course anatomy: The dissection & analysis of knowledge through teaching », in P. Hutchings (éd.), *The course portfolio: How faculty can improve their teaching to advance practice and improve student learning*. Washington, D.C: American Association of Higher Educ, 1998, p. 6.

67 À ce sujet, voir par exemple le site de l'Université de Vanderbilt (https://cft.vanderbilt.edu/guides-sub-pages/teaching-statements/).

68 http://faculty.virginia.edu/marva/Teaching%20Portfolio/teaching_portfolio.htm.

69 http://www.bu.edu/ctl/teaching-resources/documenting-your-teaching/.

70 http://as.virginia.edu/faculty-annual-review-and-post-tenure-review .

71 http://people.virginia.edu/~drp/teaching.php .

72 https://teaching.utoronto.ca/wp-content/uploads/2016/09/FMT-Report-Sept-28-2016.pdf.

IV.2 - Quelles dérives pour ces évaluations ?

IV.2.1 - Une culture de la performance et non du savoir

IV.2.1.1 - Une « réingénieurie » permanente : intégration continue des objectifs et augmentation du rendement

Dès 1998, l'AAUP avait identifié le risque majeur de ces évaluations managériales, qui élargissent les objectifs de la responsabilisation au-delà de la réddition des comptes, pour intégrer l'amélioration, les attentes et la conformité imposée par les gouvernements[73]. Ainsi leur réponse, déjà entrevue, à la mise en place du protocole de Post-Tenure Review renvoie au risque de redéfinition des profils individuels qui transparaissait plus haut chez William Locke, par des ressources humaines axées sur les objectifs d'établissement : « Many proponents of post-tenure review […] sometimes envision such review as a means for achieving larger management objectives such as "downsizing," "restructuring," or "reengineering." Individual faculty reviews should, however, focus on the quality of the faculty member's work and not on such larger considerations as programmatic direction[74] ». Les *Post Tenure Reviews* – incluant les évaluations pédagogiques – expertisent ainsi non l'exercice consciencieux et compétent d'une profession, mais l'accroissement continu de performances sans cesse réévaluées et redéfinies par les gouvernances.

Dans la réglementation de California State University en 2000, les membres titulaires sont sommés pour chaque évaluation de « repenser leurs objectifs professionnels[75] ». À l'Université du Texas, l'évaluation doit conduire les plus performants à l'être plus encore (« As a result, post-tenure reviewshould help productive faculty become even more productive[76] »). À l'Université du Kentucky, « Post-tenure review also allows for significant change in the types of work that tenured faculty do over a career, including nonvoluntary changes[77] ». On est ici devant une volonté de « choc culturel[78] » qui doit faire passer les évaluations d'une expertise centrée sur le travail de l'individu à un système proactif et axé sur l'intégration des objectifs des établissements, transformant à son gré les tâches et les objectifs assignés à chaque employé. Qu'est-ce d'autre, si ce n'est une forme d'intégration de la contractualisation sur des objectifs à court ou moyen termes, appliquée à l'ensemble des personnels, ouvrant la voie à l'orientation autoritaire des individus vers les meilleurs profils qui serviront « l'identité » de l'établissement ?

---

73 Sur cette analyse, voir Theresa SHANAHAN, « Accountability initiatives in higher education: An overview of the impetus to accountability, its expressions and implications », *Accounting or Accountability in Higher Education, proceedings of the OCUFA (Ontario Confederation of University Faculty Associations*, 2009 (en ligne), et tout particulièrement son approche des travaux de Grosjean et *al.* sur le « modèle de conformité », qui analysent la manière dont les évaluations naturalisent les objectifs des gouvernances (G. GROSJEAN,J. ATKINSON-GROSJEAN, K. RUBENSON, D. FISHER, D., *Measuring the Unmeasurable: Paradoxes of Accountability and the impacts of Performance Indicators on Liberal Education in Canada*, Ottawa: Canadian Federation for the Humanities and Social Sciences, 2000.

74 https://www.aaup.org/report/post-tenure-review-aaup-response

75 *Policy on Evaluation Of Tenured Faculty*,citée par Kelly S. JANOUSEK et Wayne DICK, « The Benefits of Pilot Testing : Post-Tenure Review at California State University, Long Beach », in Christine M. Licata, Joseph C. Morreale (éd.), *op. cit*, p. 90.

76 Debra P. PRICE, Dennis LONGMIRE, Frank FAIR, Laverne WARNER, Paul R. REED, William FLEMING, Jo Ann M. DUFFY, art. cit., p. 203.

77 James L. APPLEGATE, Lois M. NORA, art. cit., p. 39.

78 « A Clash of Cultures », *Ibid.*

Évaluation des performances et fins de carrière : un enjeu pour les SPE du CNRS ?

L'injonction d'amélioration des performances tout au long de la vie, en soi, est déjà irréaliste. En France, les seuls chiffres du suivi de carrière des chercheurs du CNRS devraient faire méditer. Le SNCS, dans un texte récent, fait remarquer que « sur les 307 chercheurs en SPE [suivi post évalution], 45% ont reçu un avis favorable lors de leur dernière évaluation, ce qui témoigne de leur retour vers une situation normale. Un point d'inquiétude est cependant que la proportion de chercheurs en SPE croît régulièrement avec l'âge : 3,7% des 50-54 ans sont en SPE, mais 4,7% des 55-59 ans et 5,4% des 60-64 ans. Au total, les plus de 50 ans forment plus de 70% des chercheurs en SPE et les chercheurs en SPE sont, à plus de 80%, des chargés de recherche. Le CNRS sait-il accompagner correctement ses chercheurs en fin de carrière ?[79] »

IV.2.2 - Des critères d'évaluation subordonnés aux attentes du marché

La question n'est pas de remettre en cause le devoir que les universités ont d'être « utiles », ni de remettre en cause leurs liens avec le tissu socio-économique des territoires, pas plus que « l'employabilité » des étudiants formés. La question est de ne pas subordonner toute l'évaluation de l'université aux attentes du marché.

Theresa Shanahan a montré comment la définition de la « qualité » de l'enseignement relevait de choix politiques. En Ontario par exemple, les dispositifs d'évaluation récompensent la conformité aux objectifs du gouvernement, alignés sur ceux de l'industrie ou du secteur privé. L'enseignement d'une discipline doit se justifier par sa capacité d'adaptation à l'économie, réorganisant les priorités universitaires en fonction de l'industrie[80]. Le lien faussement naturel établi entre « excellence », « qualité », « responsabilité » et « évaluation », renferme un piège argumentatif redoutable : ce sont des termes avec lesquels personne ne peut être en désaccord, de sorte qu'ils « organisent le consentement » jusqu'à ce qu'ils soient déconstruits lorsqu'il est révélé qu'il n'y a pas de conception commune de ce qu'ils signifient[81]. L'autre pendant du piège rhétorique repose sur l'idéologie de la concurrence, imposant le comparable là où une université créatrice devrait produire de l'incomparable. L'hypothèse douteuse qui sous-tend cette approche de la reddition de comptes est que la concurrence améliore les pratiques pédagogiques, mène à une meilleure recherche, accroît la productivité des individus, augmente l'employabilité des étudiants. Sur le fond, cette approche s'appuie sur une vision technique de l'éducation, prise dans une compétitivité économique mondiale, plutôt que sur une vision transformatrice au cœur de la démocratie.

Ce système s'adosse à des dispositifs de visibilité comme les classements internationaux, et le développement du benchmarking. Laure Endrizzi montre que si la recherche est lontemps restée prioritaire dans ces classements (comme à l'ARWU), l'enseignement est aujourd'hui pris en compte par de plus en plus de classements internationaux (comme THE World University Ranking) ou nationaux, comme le *America's Best Colleges*[82], réalisé par *US News and World Report*. De même, en Europe, on se dirige vers des classements multidimensionnels, comme la plateforme U-Multi-rank[83] (UMR) initiée par la Commission européenne en 2008 et inaugurée en 2014 avec 850 universités dont 62% européennes. Conçu par un consortium indépendant il propose notamment une évaluation prenant appui sur les résultats des évaluations par les étudiants[84].

79 http://sncs.fr/sites/sncs.fr/IMG/pdf/sncs-hebdo_18_n5.pdf
80 Theresa SHANAHAN, art. cit.
81 *Ibid.*, p. 10.
82 http://colleges.usnews.rankingsandreviews.com/best-colleges
83 http://www.u-multirank.eu/
84 Laure ENDRIZZI, art. cit., p. 8-9.

Les labellisations se situent dans une logique proche. Les labels qualité accordés aux formations, notamment d'écoles de commerce et de management, sont autant de critères d'attractivité que les institutions veulent obtenir. C'est le cas du label AACSB, fondé en 1916, qui évalue la qualité d'une Business School et compte aujourd'hui plus de 780 institutions membres accréditées à travers le monde au terme de plusieurs rapports et audits d'experts incluant la démarche pédagogique[85]. L'enquête d'un groupe de travail réuni dans une petite école de commerce accréditée par l'AACSB montre qu'il y a un lien direct entre la recherche d'obtention de ces accréditations et les facteurs d'évaluation de la pédagogie[86]. L'AACSB valorise par exemple très fortement l'innovation pédagogique. En Europe, les cadres induits par le Processus de Bologne, la perspective d'un « espace européen de l'enseignement supérieur » (EEES) et la mise en œuvre du Cadre européen des certifications ont accéléré la mesure et la mise en concurrence de la « qualité » sur des indicateurs de « performance ». Les pays signataires du Processus de Bologne doivent faire adhérer les universités à l'idée d'« assurance- qualité », susceptible d'autoriser les comparaisons et de garantir le niveau des formations (ENQA, 2009 ; ESG, 2005 ; RNCP…)[87].

IV22 - L'engagement pédagogique « observé » entre récompenses et sanctions : le dévoiement sommatif d'un outil formatif

IV221 - Quand mes pairs me jugent, ou le recul du bénéfice réel du « Peer Review of Teaching »

L'objectif premier de l'évaluation, s'il consiste en l'amélioration de la pédagogie, suppose un retour sur sa pratique et des « plans de perfectionnement ». Mais ces plans de perfectionnement, tout comme les évaluations elles-mêmes, peuvent participer d'une démarche formative, ou conditionner des décisions de carrière. À ce jour, les pratiques universitaires sont hétérogènes, mais le souci de mutualisation – économie de temps et d'argent – travaille en faveur d'une seule et même évaluation. À British Columbia par exemple, la mutualisation des évaluations est acquise pour les prises de décisions concernant le renouvellement, la titularisation, la promotion, l'avancement, le mérite…[88].

Cette confusion n'empiète-t-elle pas sur les possibilités rétroactives réelles des dispositifs formatifs ? Le rapport publié en 2013 sur le *Peer Observation and Review of teaching and Learning* par l'*Academic Professional Development Framework* de Christ Church (Canterbury) rappelle que l'observation est plus utile et acceptable lorsqu'elle est utilisée de façon formative[89]. L'évaluation sommative peut même atteindre un résultat contraire à l'objectif déclaré, en éliminant par exemple la prise de risque liée à l'expérimentation méthodologique. Dans son rapport publié en 2017, le groupe de travail de Brockport College s'inquiète de ces évaluations rédutrices : « le corps professoral ne devrait pas être puni pour des stratégies d'enseignement innovatrices qui pourraient ne pas réussir dans leurs cours[90] ».

85 https://www.letudiant.fr/etudes/ecole-de-commerce/les-labels-des-ecoles-de-commerce-19662/le-label-americain-aacsb-decrypte-14904.html D'autres labels existent, comme le label EQUIS.

86 Diane F. BAKER, Walter P. NEELY, Penelope J. PRENSHAW, Patrick A. TAYLOR, « Developing a Multi-Dimensional Evaluation Framework for Faculty Teaching and Service Performance », *Journal of Academic Administration in Higher Education*, vol. 11, iss. 2, 2015, p. 29-41.

87 Laure ENDRIZZI, art. cit., p. 5.

88 http://ore.educ.ubc.ca/files/2013/03/UBC-Principles-of-Summative-Peer-Review.pdf, p. 3.

89 Fullerton, 1999, Shortland, (2004).

90 https://brockport.edu/support/college_senate/resolutions/2013-2014%20Resolutions/RES%2039.pdf.

IV222 - Développer les primes au mérite : une roue de secours inégalitaire pour gérer la masse salariale ?

Dans des universités américaines, les primes sont perçues, et parfois même légitimées, comme les roues de secours de rémunérations très inégales que les difficultés économiques ont encore détériorées[91].

Le cas de l'Université de Washington illustre bien les mécanismes de cette détérioration et la place qu'y prennent les primes. La réglementation y stipule que des augmentations salariales peuvent survenir dans le cadre d'une promotion, mais aussi à la suite d'un ajustement salarial au mérite ou en réponse à une offre concurrentielle. La baisse des crédits de l'État, comme l'essor du dispositif de l'offre concurrentielle, sont à l'origine d'une problématique de compression, d'inversion et d'iniquité salariales qui peut placer les rémunérations de professeurs débutants rapidement au-dessus de ceux de certains collègues, qui pourtant ne déméritent pas dans leur travail quotidien[92]. Les universités tentent alors d'utiliser les ajustements salariaux au mérite pour réduire ces inégalités. Primes et bourses tentent de compenser une situation globale de pénurie, comme le reconnaît le rapport du *Committee on Faculty Responsiblities and Rewards* de l'Université de Washington: « the only useful actions that our Committee can suggest to improve upon faculty rewards are small steps to recognize special achievements or special needs of faculty in a number of different situations[93] ». Juste après la section *salaries*, le second item du document, *Rewarding Teaching Excellence*, insiste pour instituer les primes pédagogiques comme premier complément aux rémunérations, avant les bourses attribuées sur des besoins particuliers. Les recommandations sont de dépasser le statut actuel de la prime pédagogique qu'il reconnaît « modeste », « unique » et finalement « symbolique ».

Dans un rapport du printemps 2017 rendu par le Collège de Brockport (collège extrêmement sélectif et bien classé de la State University of New York), il apparaît que la substitution des primes aux augmentations salariales – certes elles-mêmes attribuées au mérite – , a aggravé la situation des personnels : « While there are other rewards that may improve faculty morale, 76% of the faculty that responded to the survey indicated that a salary increase was the most desirable reward[94] ». L'urgence est au relèvement général des progressions de carrière, notamment aux augmentations de salaire lors des promotions[95]. De plus, les enseignants rappellent que l'excellence de l'enseignement n'est pas d'abord liée à l'évaluation et aux primes mais aux conditions d'exercices du métier : et le rapport d'égréner la litanie des revendications de bon sens (éviter la surcharge des services, ne pas décider arbitrairement d'imposer des enseignements hybrides ou en ligne et tenir compte de leur poids dans les charges de travail, établir des effectifs raisonnables dans les classes…).

Les *Post Tenures Reviews* ont accentué ces inégalités. À l'Université du Nevada le réglement de l'université prévoit qu'un professeur permanent peut être congédié s'il reçoit une cote globale insatisfaisante pendant deux années consécutives[96] ; il en va de même en Virginie si

91 Alinéa 7 des recommandations, p. 59. http://www.washington.edu/reports/faculty/recommendations4.html

92 http://www.washington.edu/reports/faculty/recommendations4.html.

93 *Report of the Provost's ad hoc Committee on Faculty Responsiblities and Rewards*, p. 7, sur 174 sondés.

94 https://www.brockport.edu/academics/academic_affairs/docs/frrrp_final_report.pdf.

95 "We recommend raising the salary increase that accompanies a promotion to at least $5,000 for Assistant Professor/ Senior Assistant to Associate and at least $7,000 for Associate to Professor", p. 54 (les salaires étant calculés sur une année universitaire de 9 mois). Dans ce même rapport, les enseignants demandent à être rémunérés – ne serait- ce qu'en décharge – et inversement que l'allocation des présidents ne dépassent pas 10% du salaire de base pendant l'année académique et pendant l'été…

96 Donna R. RUBEN, Barbara LEE, « Managing Faculty Productivity After Tenure », *Legal Issues in Higher Education*, 2005 (https://www.aaup.org/issues/tenure/managing-faculty-productivity-after-tenure-2005).

l'enseignant n'atteint pas au bout d'un an les objectifs énoncés par son plan de perfectionnement[97] ; dans la réglementation de Indiana University - Purdue University Indianapolis, imposée au début des années 2000, les sanctions vont du gel des salaires au congédiement, en passant par la perte de l'indemnité de déplacement, la perte de l'admissibilité au congé sabbatique, la rétrogradation des fonctions, la pension spéciale sans solde, et la réduction du salaire, concept sans précédent à l'IUPUI.

IV.2.3 - La machinerie de l'évaluation : un sous-investissement dans l'humain

IV.2.3.1 - La machine à évaluer : les pairs sont-ils des pairs, les soutiens sont-ils des soutiens ?

L'accumulation d'évaluations « multifactorielles » aux sources emboitées (pair-mentor, pair-évaluateur, étudiants, experts, représentants de la gouvernance…) transforme le dossier d'un universitaire en un gigantesque ensemble d'indicateurs mesurables et exploitables, faisant des universités des machines à produire de l'évaluation au nom de la qualité.

Si dans l'observation entre pairs à visée formative, voire dans le *mentoring*, peut s'établir un échange collégial et salutaire, les protagonistes de l'évaluation sommative ne sont pas exactement dans une situation d'égalité. Les pairs évaluateurs apparaissent de plus en plus comme des experts au sein d'une élite aux contours largement autodéfinis. L'identification des « mentors », par exemple, pose un réel problème. Le *Guidebook* de l'Université d'Austin souligne que Katz et Henry (1988) ont fait la promotion d'un programme de maîtrise qui jumelait un professeur chevronné et un professeur subalterne, mais qu'on pourra préférer des comités de deux, trois ou quatre évaluateurs aux simples binômes[98]. Le même document envisage aussi la possibilité de faire intervenir des étudiants diplômés, comme à la *Graduate School of Business* de l'Université de Chicago[99]. Une chose est certaine : l'ampleur de la machine crée une nouvelle catégorie d'universitaires, experts non tant en pédagogie –qui relève des sciences de l'éducation – qu'en évaluation de la pédagogie. Pour l'Université de Toronto, l'activité de mentoring doit devenir une activité universitaire à part entière. Voie de promotion, elle a vocation à constituer une élite de la pédagogie au sein des disciplines, filière de mentors capables de former de nouveaux « champions » (sic[100]). Ces champions sont, à Toronto, potentiellement distingués par le « university-wide President's Teaching Award », qui donne accès, pour une période minimale de cinq ans, à « l'Académie d'enseignement », qui discute régulièrement des questions relatives à l'enseignement à l'Université, donne des conseils à la gouvernance et au directeur du *Center for Teaching Support & Innovation* (CTSI) et … aide à l'évaluation de l'enseignement[101]. S'impose la figure des « pairs-évaluateurs » comme nouvelle élite[102].

Aux côtés des experts, les conseillers pédagogiques des centres d'innovation à la pédagogie, créés dans toutes les universités, jouent aussi un role clé, et parfois ambigu. Soutien précieux de la formation, ils sont des acteurs majeurs de cette « gouvernance rationalisée » définie en Europe

97 Susan H. BARR, art. cit., p. 209.

98 http://www.tlataskforce.uconn.edu/docs/resources/Texas_Peer_Observation_Guidelines.pdf, p. 6. Voir J. KATZ, M. HENRY « Turning Professors into Teachers: A New Approach to Faculty Development and Student Learning », New York: Macmillan, 1988.

99 *Ibid.*, p. 7.

100 https://teaching.utoronto.ca/wp-content/uploads/2016/09/FMT-Report-Sept-28-2016.pdf, p. 11-19.

101 http://www.provost.utoronto.ca/awards/ presidentaward/about_pta.htm

102 Christine MUSSELIN, « 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, 2013, p. 1165-1173, citée par Laure Endrizzi, art. cit., p.11.

par les travaux du *High Level Group on the Modernisation of Higher Education* (2013) pour la Commission européenne. L'enseignant-chercheur partage la responsabilité de ses activités avec les équipes dirigeantes, les collègues, les étudiants, et désormais les « cellules qualité ou ou structures d'appui avec un mandat explicite[103] ». De la conception des instruments pédagogiques à la participation active à l'évaluation, en passant par la formation et la remédiation, ils sont à la fois conseillers et experts. Ne serait-ce que par le poids des disciplines dans les approches pédagogiques, il y a là un équilibre à trouver, les experts de la pédagogie ne pouvant se substituer aux pédagogues de leur discipline.

IV232 - La prolifération de dispositifs chronophages à l'efficacité discutable

Theresa Shanahan rappelle combien ces dispositifs alourdissent le fardeau du corps professoral sans nécessairement en améliorer la qualité. Pour Janice Currie cette « bureaucratie mesquine et le remplissage de formulaires de plus en plus nombreux » ne fait que modifier le processus de travail des universitaires et accroître les exigences administratives du corps professoral[104]. Ces pratiques restructurent le travail quotidien, le culte de la responsabilité s'épanouissant sous la forme d'audits, d'exercices de mesure, de classements des écoles sur la « performance », de tests standardisés et d'accréditation[105].

Parcourir les universités anglo-américaines, c'est voir se multiplier à l'envie les différents protocoles de *Peer Reviews*, dont l'organisation effective pèse souvent sur des départements déjà exsangues. À l'université de Takoma[106], la périodicité des évaluations, et les données qu'elles doivent contenir, varient d'un département à l'autre, mais la réglementation envisage très sérieusement l'intérêt d'une évaluation hebdomadaire. Sur le campus de British Columbia, l'évaluation par les pairs, mêlant le *mentoring* et l'évaluation sommative, génère des cascades d'évaluations. Chaque département organise le *Peer Review* à sa guise, pour peu qu'il suive le cahier des charges de l'université : sont mobilisés un pair « mentor », un pair évaluateur, des examinateurs indépendants pour étudier les biais contextuels, un examinateur externe pour rédiger le rapport à partir des commentaires de chacun après une approche en équipe, et avant la relecture collégiale du rapport… Chaque évaluation revient presque à la somme de travail d'un comité HCERES visitant une unité, pour un seul enseignant-chercheur. Les départements se voient confier la mise en œuvre des lignes directrices de l'université « en tenant compte de la culture et des contextes d'enseignement de l'unité (faculté/école/département/département), [et] de la discipline (le type de cours, la pédagogie pertinente à la discipline, les modalités de l'enseignement, etc) ». L'université d'Austin affirme qu'au moins trois observations en classe au cours d'un semestre ou d'un trimestre sont nécessaires, mais reconnaît aussi qu'il faudra des « rotations annuelles » car « les petits départements auront de la difficulté à cet égard[107] ». Dans le même temps, l'Université de Toronto envisage des dispositifs d'évaluation pour évaluer les dispositifs d'évaluation[108]…

103 Laure ENDRIZZI, art. cit., p. 10.

104 Janice CURRIE, « Introduction», in CURRIE Janice et NEWSON Janice K. (éd.), *Universities and globalization: Critical perspectives*, Thousand Oaks, CA: Sage 1998, p.1-13, cité par Shanahan, art. cit., p. 13.

105 Shanahan rappelle que dans *The Evaluative State Reconsidered* (1998), Guy Neave les décrit comme des « instruments d'une nouvelle surveillance » et a qualifié la responsabilisation dans l'enseignement supérieur de « véritable orgie » de procédures de rapport (p. 266.)

106 http://www.tacoma.uw.edu/sites/default/files/sections/FacultyAssembly/Evaluating_Teaching_P.T._Guide_Draft_Kalikoff_TLC_UWS.pdf. Le rapport date de 2016.

107 https://ctlt.ubc.ca/programs/all-our-programs/peer-review-of-teaching-2/.

108 https://teaching.utoronto.ca/wp-content/uploads/2016/09/FMT-Report-Sept-28-2016.pdf.

IV233 - Entre mesure, objectivation et surmenage : peut-on intégrer sérieusement les biais contextuels ?

Quantifier et/ou décrire des pratiques et des ressentis rapportés à des critères communs et comparables ne permet pas d'inclure aisément les contextes pédagogiques, liés au parcours de l'enseignant, au type d'enseignement (discipline, filière, cadre pédagogique…), et au public concerné (niveau, effectifs, motivation…), qui doivent pondérer l'usage concurrentiel de ces évaluations entre collègues, formations et universités.

Les recherches qui se développent en faveur d'une évaluation systémique de l'enseignement sont certes conscientes de la complexité de ces facteurs. Pour Graham Gibbs (2010) le modèle des « 3P » (presage, process, product) de Biggs[109] permet d'intégrer les éléments contextuels propres à l'établissement (qualité du public accueilli, formation des enseignants, nature et poids de la recherche…), les critères relatifs au processus d'enseignement (effectifs des classes, niveau d'engagement des étudiants, nature et quantité du feed-back reçu sur leur travail etc…) et les critères relatifs aux résultats (acquis d'apprentissage, diplomation, employabilité etc…). Le problème est qu'on aboutit alors à des possibilités de modélisations multiples, en fonction des éléments valorisés. Les biais contextuels (*presage*) pèsent généralement peu, et c'est au profit du *process* et du *product* que les oppositions se jouent[110].

Certes, dans un protocole d'observation sérieux entre pairs, le contexte est toujours pris en compte : l'observation doit être précédée d'un « Peer-to-Peer Interview » incluant « les antécédents personnels et professionnels de son collègue et ses réflexions sur l'enseignement[111] ». Mais, comme l'écrivent Cara Gormally et *al*., l'élément individuel est difficile à objectiver, ne serait-ce que par le manque d'expertise du corps professoral en matière de rétroaction pédagogique, le parti pris des observateurs envers un style d'enseignement semblable, les problèmes de fiabilité ou les conflits d'intérêts…

IV234 - Les effets pervers du sous investissement face aux enjeux

En 2002, seuls 36 % des établissements affectaient explicitement des fonds pour aider les membres du corps professoral à réaliser leurs plans, alors que dans 38% des établissements des sanctions pouvaient déjà être imposées après le *Post-Tenure Review*[112]. L'évaluation pédagogique joue un rôle clé dans les décisions de carrière, alors que la formation à la pédagogie n'est toujours pas une vraie priorité, même si les partisans d'une institutionnalisation radicale du SoLT l'érigent comme le meilleur modèle de formation et de recherche sur la pédagogie – ce qui mérite au moins débat avant d'être importé tel quel en France, du moins eût égard aux débats qu'il suscite outre-atlantique. Même l'Université du Kentucky, où le Post-Tenure Review a été accompagné de la création d'un fonds de formation professorale d'un million de dollars par le gouvernement de l'État, reconnaît que trop souvent l'évaluation s'appuie sur les sanctions comme facteurs de motivation, plutôt que sur la promesse d'une « bourse » significative de formation ou d'expérimentation pédagogique[113].

Si la question de la formation des enseignants est posée, en amont et en aval de l'évaluation, celle des observateurs/évaluateurs l'est plus encore. En l'état actuel des choses, l'insuffisance de formation des universitaires discrédite plutôt le dispositif de l'observation entre pairs, tout

109 J. B. BIGGS, « From Theory to Practive : a Cognitive Systems Approach », *Higher Education Research and Development*, 12 (1), 1993, p. 73-85. Voir notamment ici Laure ENDRIZZI, art. cit., p. 12.
110 Ibid.
111 Erica CAMPBELL, « The value of the non-evaluative: Rethinking faculty observation », art. cit.
112 Cathy A. TROWER, art. cit., p. 56.
113 James L. APPLEGATE, Lois M. NORA, art. cit.

particulièrement dans sa dimension sommative. La récurrence des témoignages en ce sens doit nous mettre en garde sur la capacité du courant SoTL à véritablement former à la pédagogie des universitaires déjà soumis à l'explosion des tâches. Dans son article « Don't tell it like it is », Isabeau Iqbal s'appuie sur trente entretiens avec des professeurs d'universités canadiennes, issus de la Faculté des Sciences et de la Faculté des arts de l'Université of Western Canada[114]. Certains ont été évalués et évaluateurs, 11 ont été évalués, 10 ont été évaluateurs. L'écart avec la littérature apologétique des rapports justificatifs de la pratique est alors frappant. Presque tous soulignent que la rétroaction des évaluations sommatives n'a pas contribué à leurs progrès professionnels. Ils déplorent les limites de l'observation, portée sur le comportement du professeur et sur la réaction des étudiants (sont-ils attentifs, distraits, le professeur parle-t-il assez fort…) plutôt que sur les questions d'apprentissage elles-mêmes. Ces enseignants rappellent d'ailleurs qu'ils ne sont pas vraiment formés à la pédagogie, ni à l'évaluation de la pédagogie[115]. Ils soulignent aussi que les observations sont des situations d'enseignement artificielles, reposant sur « une petite partie des activités pédagogiques du collègue ».

Présent sur de nombreux sites universitaires, l'article d'Erica Cambpell « The Value of the Non-Evaluative: Rethinking Faculty Observation » (2017) argumente pour un démantèlement des structures d'évaluation sommative au profit de l'observation formative articulée à un système de « récompenses » non par de simples primes mais par des bourses, semestres sabbatiques, aménagements de service pour s'engager davantage dans une réflexion pédagogique choisie. Elle propose un dispositif pensé pour privilégier la croissance professionnelle plutôt que le jugement de valeur et un outil de gouvernance. Une telle ambition doit aboutir à la valorisation du travail en équipe. Pour Campbell, les départements doivent organiser de vraies équipes pédagogiques échangeant sans arrière-pensées.

Très présentes également sur les sites, Cara Gormally, Mara Evans et Peggy Brickman vont plus loin en soutenant que des modèles de rétroaction ou d'encadrement par les pairs plutôt que d'observation et d'examen par les pairs pourraient encourager davantage l'utilisation efficace de stratégies d'enseignement. Elles aussi soulignent la supériorité des équipes pédagogiques sur le face-à-face évaluateurs/évalués. Réfléchissant à l'enseignement des sciences, elles montrent également, c'est suffisamment rare et important pour être souligné, à quel point il est impératif de disposer de centres de formation étudiant les pratiques pédagogiques de manière ciblée, par disciplines et par champs, comme le proposent pour l'enseignement des sciences *the Faculty Institutes for Reforming Science Teaching*[116].

Ainsi, au lieu de servir à la gestion du personnel, le feed-back devrait se faire à l'échelle de la communauté, « publiant » pour tous les retours d'expérience : méthodes d'enseignement, formation d'ateliers de perfectionnement professionnel, recueils de défis et de questions, publication de rapports fondés sur les résultats collectifs de l'auto-réflexion. Allant au bout du processus Cambpbell suggère la possibilité de l'anonymat pour les enseignants volontaires qui souhaiteraient cependant ne pas s'identifier dans ce partage collectif d'expérience.

114 Isabeau IQBAL, « Don't tell it like it is: Preserving collegiality in the summative peer review of teaching », *Canadian Journal of Higher Education Revue canadienne d'enseignement supérieur*, 44, 1, 2014, p. 108 -124.

115 Ainsi de ce professeur questionné par Iqbal et qui répond : « "We don't know about pedagogy and don't have training as teachers. So, there is a reticence about telling people how to do their job", *op. cit.*, p. 115.

116 Cara GORMALLY, Mara EVANS, Peggy BRICKMAN, art. cit., p. 193.

Un exemple « pionner » de prime d'engagement pédagogique : la PEP d'Aix-Marseille Université

Depuis 2016, Aix-Marseille Université expérimente la PEP, Prime d'Engagement pédagogique. Cette prime, contrairement à la PEDR, a un statut de prime d'intéressement en accord avec la circulaire ministérielle DGRH A1-2 n° 0023 du 17 février 2017[117] qui précise les modalités de création de régimes d'intéressement sur le fondement de l'article L954-2 du code de l'éducation au sein des établissements publics d'enseignement supérieur ayant accédé aux responsabilités et compétences élargies. Elle ne peut être transformée en décharge et attribue une somme annuelle de 2000 euros non renouvelable d'une année sur l'autre (voir aussi encadré suivant sur l'Université de Strasbourg).

Les critères d'évaluation reposent sur les 8 compétences fonctionnelles du référentiel de compétences pédagogiques universitaires d'AMU (Concevoir un enseignement - Utiliser les différentes Technologies de l'Information et de la Communication - Transmettre des savoirs universitaires - Encadrer et accompagner l'étudiant - Evaluer les apprentissages - Travailler en équipe - Coordonner un programme pédagogique universitaire - Réfléchir à sa pratique pédagogique et la faire évoluer). Les primes sont validées en CAC restreint, d'après l'avis de chaque composante et du CIPE, le Centre d'Innovation Pédagogique et d'Évaluation. Bien que la décision finale revienne au CAC restreint, l'avis du CIPE, chargé de l'expertise approfondie de chaque dossier, est prépondérant dans les attributions. Il en ressort une attention soutenue aux innovations pédagogiques – classe inversée, *serious games*… – et aux supports numériques, qui permettent notamment l'enseignement non présentiel. La prime à l'intéressement relève bien ici du nouveau management public, qui a pour objectif de soumettre la communauté aux objectifs d'établissement et aux réformes en cours[118]. Ces critères permettent-ils de prendre en considération les spécificités de toutes les filières et toutes les disciplines ? De nombreux lauréats, enseignent dans des filières d'excellence ou à des niveaux élevés (master, doctorat). En L1, la responsabilité d'équipes est un critère valorisant, mais qui occulte précisément la cohérence et le travail de l'équipe en récompensant un seul leader. La prime ne reconnaît elle-même l'engagement que d'une extrême minorité, une trentaine de lauréats par an, 29 en 2018, dans une université qui compte près de 2800 enseignants titulaires, et près de 1500 enseignants contractuels, souvent très investis, qui ne participent pas au dispositif.

Prime individuelle *vs.* prime collective : la réponse de la liste « Alternative 2017 » à la PRIP mise en place à l'Université de Strasbourg[119]

Le CA de l'Université de Strasbourg consacrait le 13 mars 2018 l'arrivée de la PRIP, (« Prime de Reconnaissance de l'Investissement Pédagogique »). À ce dispositif attribuant des récompenses élevées (8000 euros sur deux ans) à un nombre très restreint d'enseignants-chercheurs (15 lauréats, moins de 3 personnels pour mille, par an) récompensés par une procédure discrétionnaire, la liste Alternative 2017 proposait de reconnaître avant toute chose que « l'enseignement est une activité éminemment collective ». C'est de l'équipe pédagogique que dépend la formation cohérente des étudiants. C'est l'équipe pédagogique, et non le prédicat vide de sens « qu'*il existe toujours un leader* », qu'il faut encourager, mutualisant les efforts et soudant les individus au lieu de les fragiliser par une vaine concurrence. Et *via* l'équipe pédagogique, c'est l'ensemble de la communauté qu'il faut construire par la publication des expérimentations réussies : non des dossiers confidentiels opposés les uns aux autres, mais une sommes de savoirs collégialement construits et transmis.

117 http://alternative2017.eu/wp-content/uploads/2017/11/04b-Circulaire-minist.-DGRH-A1-2-0023-du-17.02.2017.pdf.

118 Voir ici les analyses de Theresa Shanahan (art. cit.) à partir des analyses de Grosjean et *al.*, op. cit. sur le « modèle de conformité » de la responsabilisation, qui intègre les priorités de l'établissement aux bonnes pratiques sanctionnées positivement par les évaluations.

119 http://alternative2017.eu/index.php/category/elus/.

Comme beaucoup de chercheurs spécialistes des questions d'évaluation, Laure Endrizzi constate que « la question de l'estime de soi dans le cadre des EEE est très insuffisamment traitée par la recherche, [et] 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 ou moins longtemps quand il est négatif[120] ». Cet impact désastreux des évaluations négatives, incapables de modifier sur le long terme les pratiques de l'enseignant voire les aggravant, est actuellement questionné par la recherche anglo-américaine. Pour Gormally et *al*., les pratiques courantes de rétroaction pédagogique dans l'enseignement supérieur ne tiennent pas compte des personnalités mises en jeu, et les membres du corps professoral reçoivent généralement une rétroaction négative, ou ce que Utell (2013) appelle une « rétroaction fondée sur l'échec », soulignant les défaillances.

Par ailleurs, de plus en plus d'enquêtes reconnaissent l'importance de la souffrance générée par une pression absolument constante d'évaluations. L'énorme rapport produit par le *Center for Teaching Support and Innovation* de l'Université de Toronto sur l'innovation à l'automne 2016 (67 pages)[121] souligne combien l'adhésion au dispositif sommatif d'évaluation par les pairs masque mal une anxiété et une insatisfaction extrêmes. Ce document intitulé *Faculty Mentoring for Teaching Report* se fonde sur une enquête de terrain auprès des personnels, et fait suite aux résultats de deux enquêtes récentes relevant des failles dans le système même de « mentoring ». Dans la première enquête (*Collaborative on Academic Careers in Higher Education* – COACHE
– *Survey*, 2012), il apparaît qu'à Toronto seuls 48% des interrogés estimaient le mentorat efficace. Le sondage « Speaking Up », réalisé en 2014 à l'Université de Toronto, estimait que plus du cinquième (21,5 %) des membres du corps professoral titulaire ou en voie de titularisation ont signalé un stress « considérable » associé à leurs responsabilités d'enseignement
– et que près de la moitié (45 %) des nouveaux membres du corps professoral ont connu des niveaux de stress « importants ». La crainte d'expérimenter de nouvelles approches, la concurrence, l'isolement associé à la nécessité de combler seul ses lacunes pour les masquer, sont autant de retombées décrites.

Le stress est d'autant plus lourd qu'il se conjugue au poids de la « réflexivité pédagogique » contrainte et permanente. De ce point de vue, le courant SoTL divise : occasion de se saisir de sa pratique pédagogique ? Tâche supplémentaire remplie d'effets pervers ? L'effet pervers par excellence apparaît dans la mutation du SoTL en injonction dans le cadre des évaluations professionnelles, ainsi que le souligne Bill Cerbin, professeur en psychologie à l'Université du Wisconsin : « Comme [...] toute recherche, le SoTL doit résulter d'un engagement. personnel. La pire motivation consiste à le faire parce que cela fait partie des tâches à accomplir en vue d'atteindre la titularisation. C'est là le meilleur moyen de faire d'une partie importante du travail une simple tâche irritante supplémentaire [122]».

120 Laure ENDRIZZI, art. cit., p. 34. Elle se réfère notamment aux travaux d'Hervé BARRAS, « 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. 2014.

121 https://teaching.utoronto.ca/wp-content/uploads/2016/09/FMT-Report-Sept-28-2016.pdf

122 Bill CERBIN, « Investigating Student Learning in a Problem-Based Psychology Course", in Pat Htuchings (éd.), *Opening Lines: Approaches to the Scholarschip of Teaching and Learning*, Carnegie Publications, the Carnegie Foundation for the Advancement of Teaching, 2000, p. 3-4, cité par Mélissa Thériault, art. cit.

De la formation à l'évaluation : les universités françaises à l'heure du SoTL ?

Publié en mai 2017, le décret relatif à la formation des maîtres de conférences a instauré une obligation d'accompagnement dès la rentrée 2018. Ils bénéficieront d'une « formation visant l'approfondissement des compétences pédagogiques nécessaires à l'exercice du métier », sous la forme d'une décharge de 32 heures durant la première année d'exercice. C'est louable, c'est un début, mais qu'approfondit-on au juste (« approfondir » laissant supposer qu'il y a bien matière à creuser) : sur quelles formations antérieures communes construit-on la formation des enseignants dont la pédagogie sera ensuite mise en concurrence lors des évaluations ? comment les spécificités disciplinaires sont-elles abordées par les conseillers pédagogiques ? Surtout, qu'approfondit-on en si peu de temps, selon quelles méthodes et pour quels objectifs, sachant que la formation est marquée par l'empreinte des « philosophies » locales ? Ainsi, l'Idip de Strasbourg[123], *Institut de développement et d'innovation pédagogiques*, revendique clairement sa filiation au courant SoTL, et propose dans son DESU l'ensemble des figures imposées de l'évaluation pédagogique outre-atlantique : portfolios réflexifs, rédaction d'articles sur son projet pédagogique associé à la construction du « projet pédagogique de transformation ou d'innovation pédagogique », tenue d'une conférence publique… Entreprises passionnantes, qui ne doivent pas ignorer, pourtant, les vifs débats en cours aux États-Unis sur le lien insidieux entre SoTL et évaluations sommatives, transformant l'engagement personnel en injonction chronophage pour tous, pourvoyeuse de « niches » pour quelques uns…

BIBLIOGRAPHIE

Évaluations et évaluations de la pédagogie

BAKER Diane F., NEELY Walter P., PRENSHAW Penelope J., TAYLOR Patrick A., « Developing a Multi-Dimensional Evaluation Framework for Faculty Teaching and Service Performance », *Journal of Academic Administration in Higher Education*, vol. 11, iss. 2, 2015, p. 29-41.

BOYER Ernest L., Scholarship reconsidered : Priorities of the Professoriate, The Carnegie Foundation for the Advancement of Teaching, 1990 (éd. augmentée par Drew MOSER, Todd C. REAM et John M. BRAXTON en 2015).

CENTRA, John A., *Reflective Faculty Evaluation*. San Francisco: Jossey-Bass Publishers, 1993.

CHAIT Richard P. (éd.), *The Questions of Tenure*, Harvard University Press, Cambridge, London, 2002. COHEN, P.S. et MCKEACHIE, W.J. (1980) "The Role of Colleagues in the Evaluation of College Teaching."
Improving College and University Teaching, 28, p. 147-154.

ENDRIZZI Laure , « 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, 2014.

GIBBS Graham, *Dimensions of Quality*, York, The Higher Education Academy, 2010.

GIBBS Graham, *Learning by Doing, A Guide to Teaching and Learning Methods*, Oxford Polytechnic, 1988. GOLPARIA Shaya, CHAN Judy, CASSIDY Alice, « Peer Review of Teaching : Sharing Best Practices »,
Collected Essays in Learning and Teaching, vol. VIII, 2015, p. 211-218.

GOSLING David, « Collaborative Peer-Supported Review of Teaching », in J. SACHS et M. PARSELL (éd.), *Peer Review of Learning and Teaching in Higher Education. International Pespectives*, Springer, 2014, p. 13-27.

KATZ, J. and Henry, *Turning Professors into Teachers: A New Approach to Faculty Development and Student Learning*. New York: Macmillan, 1988.

LICATA, Christine M., MORREALE, Joseph C. (éd.), *Post-Tenure Faculty Review and Renewal : Experienced Voice,* American Association for Higher Education, Washington, DC. 2002.

LOCKE William, « Intégrer les stratégies de recherche et d'enseignement : les implications pour la gestion et la conduite des établissements au Royaume-Uni », *Politiques et gestion de l'enseignement supérieur*, 2004/3 (nº 16), p. 115-137.

RUBEN Donna R., LEE Barbara, « Managing Faculty Productivity After Tenure », *Legal Issues in Higher Education*, 2005 (en ligne, https://www.aaup.org/issues/tenure/managing-faculty-productivity-after-tenure- 2005).

SACHS Judith et PARSELL Mitch (éd.), *Peer Review of Learning and Teaching in Higher Education. International Pespectives*, Springer, 2014.

SELDIN, Peter (1997) *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*, 2nd edition. Bolton, MA: Anker Publishing Company, Inc. Additional information about compiling a teaching portfolio, with examples, is available from the University of Massachusetts at Amherst (1ère éd. 1995)

SHULMAN, Lee S., « Course anatomy: The dissection & analysis of knowledge through teaching », in P. Hutchings (éd.), *The course portfolio: How faculty can improve their teaching to advance practice and improve student learning*. Washington, D.C: American Association of Higher Educ, 1998, p. 5-12.

TROWER Cathy A., "What Is Current Policy?", in Richard P. Chait, *The Questions of Tenure, op. cit.*, p. 32- 68.

Reculs critiques

CAMPBELL, Erica, « The value of the non-evaluative: Rethinking faculty observation », in The Graduate Center Learning Collective City University of New York and Hilarie Ashton (éd.), *Structuring equality: A handbook for student centered learning and teaching practices*, p. 109-130 https://www.hastac.org/collections/structuring-equality-handbook-student-centered-learning-and-teaching- practices

CURRIE Janice et NEWSON Janice K. (éd.), *Universities and globalization: Critical perspectives*, Thousand Oaks, CA: Sage 1998.

GOLPARIAN, Shaya, CHAN, Judy, CASSIDY, Alice, Peer *Review of Teaching: Sharing Best Practices, Collected Essays on Learning and Teaching,* 2015, p. 211-218.

GORMALLY Cara, EVANS Mara, BRICKMAN Peggy, « Feedback about Teaching in Higher Education : Neglected Opportunities to Promote Change », *Life Sciences Education,*vol. 13, 187–199, Summer 2014

GROSJEAN Garnet , ATKINSON-GROSJEAN JANET, RUBENSON Kjell, FISHER Donald, *Measuring the Unmeasurable: Paradoxes of Accountability and the impacts of Performance Indicators on Liberal Education in Canada*, Ottawa: Canadian Federation for the Humanities and Social Sciences, 2000

HUSTON Therese, WEAVER C.L. (2008), « Peer coaching: professional development for experienced faculty », *Innovative Higher Education*, 33, p. 5-20

IQBAL Isabeau, "Don't tell it like it is: Preserving collegiality in the summative peer review of teaching",
Canadian Journal of Higher Education, vol. 44, No. 1, 2014, p. 108-124

MILLER, J. Elizabeth, SELDIN Peter, « *Changing Practices in Faculty Evaluation*: *Can Better Evaluation Make* a *Difference*?», AAUP Reports and Publication. May-June 2014.

MULLER Jerry Z., *The Tyranny of Metrics*, Princeton, 2018 *MILLER, J. Elizabeth, SELDIN Peter,* « *Changing Practices in Faculty Evaluation*: *Can Better Evaluation Make* a *Difference*?», AAUP Reports and Publication. May-June 2014

SHANAHAN, Theresa, « Accountability initiatives in higher education: An overview of the impetus to accountability, its expressions and implications », Accounting or Accountability in Higher Education,

proceedings of the OCUF, 2009

VINOKUR Annie et EYRAUD Corine, « Le “Higher Education and Research Act 2017” : acte de décès du service public de l’enseignement supérieur en Angleterre ? », *Droit et société,* 2018/1, n° 98, p. 113-138.

SoTL

BIÉMAR, S., DAELE, A., MALENGREZ, D. & OGER, L., « Le « scholarship of teaching and learning » (sotl). proposition d’un cadre pour l’accompagnement des enseignants par les conseillers pédagogiques », *revue internationale de pédagogie de l’enseignement supérieur,* 31(2) (2018) http://journals.openedition.org/ripes/966