

19<sup>e</sup> Colloque International de l'ADMEE-Europe

L'évaluation au 21<sup>e</sup> siècle : Vers de nouvelles formes, modélisations et pratiques de l'évaluation ?

11-13 septembre 2006, Luxembourg

Résumés des conférences

	<b>Titre</b>	<b>Auteur</b>
Conférence d'ouverture	L'ÉVALUATION AU 21E SIÈCLE	Professeur Jean-Marie De Ketele, Université Catholique de Louvain
Conférence de clôture et débat	POUR UNE ÉPISTÉMOLOGIE DE L'ÉVALUATION	Professeur Gérard Figari, Université Pierre Mendès, Grenoble
Conférence 1	LE TESTING ASSISTÉ PAR ORDINATEUR DANS LA MESURE EN EDUCATION : POTENTIALITES ET LIMITES	Professeur Romain Martin, Université du Luxembourg
Conférence 2	L'ÉVALUATION DES ÉLÈVES : ENTRE MESURE ET JUGEMENT	Professeur Marcel Crahay, Université de Genève
Conférence 3	ÉTAT DES LIEUX ET PERSPECTIVES POUR L'ÉVALUATION DE LA QUALITÉ DE L'ENSEIGNEMENT SUPÉRIEUR EN EUROPE	Fabrice Henard, Comité National d'Évaluation, France
Conférence 4	UN NOUVEAU PARADIGME POUR LES ENQUÊTES INTERNATIONALES : D'UNE PERSPECTIVE SYNCHRONIQUE VERS UNE PERSPECTIVE DIACHRONIQUE	Christian Monseur, Université de Liège

**Conférence d'ouverture: L'évaluation au 21e siècle**

**Professeur Jean-Marie DE KETELE**, Université Catholique de Louvain

Les 20 ans d'existence de l'ADMEE se sont traduits par de nombreuses publications. Beaucoup d'entre elles ont été publiées dans les nombreux ouvrages issus des colloques organisés sous l'égide de l'association ou dans la revue « Mesure et Evaluation en Education ». Compte tenu de l'ampleur des travaux des chercheurs membres de l'ADMEE en l'espace de deux décennies, des synthèses ont déjà été effectuées par des membres particulièrement actifs. Nous pensons particulièrement à (1) la synthèse de Cardinet et Laveault publiée en 1996 sur « Dix années de travaux européens et nord-américains sur l'évaluation : quelles lignes de force » dans « Mesure et Evaluation en Education » ou à (2) celle de Figari et Achouche publiée dans la même revue en 1997 sur « Dix ans de travaux de recherche en évaluation » , ou plus récemment à (3) la revue critique de 105 publications francophones réalisée par Allal et Mottiez-Lopez en 2005 pour l'OCDE sous le titre « Formative assessment : improving learning in secondary classrooms ». Plus récemment encore et grâce au travail éditorial remarquable de Figari et Mottiez-Lopez, 26 membres actifs de l'ADMEE ont tenté de synthétiser les grandes tendances des travaux effectués dans le domaine de l'évaluation. Cette publication sera disponible en avant-première à Luxembourg.

Nous avons eu la chance de pouvoir lire ou relire ces travaux. La présente conférence n'a pas la prétention de faire une nouvelle synthèse de ces travaux, mais de dégager ce que nous avons appelé des « dimensions dynamiques » qui traversent les travaux analysés, indiquent des évolutions et permettent de dégager quelques grandes questions qui seront sans doute la base des travaux des prochaines années. A titre d'exemple, la première dimension dynamique que nous avons dégagée est celle que nous avons intitulée « De la docimologie à l'évaluation formative et de celle-ci à la notion transversale de régulation ». D'autres dimensions dynamiques seront identifiées, mais nous laissons le soin aux participants de les découvrir. C'est donc un regard rétrospectif et prospectif que la conférence va tenter de donner pour ouvrir les travaux de ce colloque anniversaire.

**Conférence de clôture : Pour une épistémologie de l'évaluation.**

**Gérard Figari**, Université Pierre Mendès France

**Introduction**

On peut toujours aborder l'épistémologie d'une discipline : le caractère organisé de cette dernière, soumis à des paradigmes consensuels peut offrir des discours construits et reconnaissables sinon consensuels sur sa valeur scientifique. Il n'est pas aussi aisé d'aborder l'épistémologie d'un champ de recherche au sein d'une discipline de par la nature heuristique des problématiques et la particularité des objets étudiés. Il est encore plus risqué de vouloir aborder l'épistémologie d'un champ d'étude dans un ensemble pluridisciplinaire comme les Sciences de l'Education. C'est le cas pour l'évaluation.

D'autant plus que, si l'on veut, aujourd'hui, faire le point sur l'évolution des savoirs dans le domaine de l'évaluation en éducation, on s'aperçoit très vite que la pensée sur cet objet s'est dispersée depuis les années 1960-1980, époque heureuse mais révolue où les grands paradigmes unificateurs et les maîtres à penser étaient universellement reconnus.

Aujourd'hui, on peut dire qu'une épistémologie de l'évaluation n'existe pas en tant que telle dans le paysage des écrits sur ce sujet,

**Alors, pourquoi aborder cette question de l'épistémologie de l'évaluation ? Pourquoi maintenant ? Pourquoi ici ? Comment l'aborder ?**

*Pourquoi aborder cette question ?*

Parce qu'il semble difficile de justifier qu'un champ de recherche comme le nôtre puisse vivre et se développer sans rencontrer le questionnement épistémologique ! Comment imaginer, en effet, que nous ne posions pas des interrogations sur le statut des résultats que l'évaluation produit ? Par exemple, ces résultats constituent-ils des jugements éphémères et discutables, des "savoirs locaux", des connaissances sur un type de phénomènes comme les effets des apprentissages réalisés dans telles ou telles conditions, les difficultés et les échecs des apprenants, la qualité des systèmes d'enseignement et de formation ou encore des connaissances sur les processus d'évaluation eux-mêmes ? Malgré l'absence de textes explicites sur le statut épistémologique de l'évaluation, on trouve des réponses à ces interrogations, ici et là, dans les publications et communications qui traitent notre sujet. Organiser les conditions d'une confrontation des postures épistémologiques des chercheurs en évaluation ne pourrait que clarifier et faire progresser les travaux qui en relèvent.

*Pourquoi maintenant ?*

Les pratiques et les dispositifs d'évaluation se sont particulièrement amplifiés et rencontrent suffisamment de critiques et de remises en question pour susciter, aujourd'hui, plus que jamais, un besoin de justification scientifique pour une meilleure reconnaissance dans le monde universitaire, professionnel et social.

*Pourquoi ici ?*

Existe-t-il un meilleur lieu que l'ADMEE et une meilleure occasion que ce colloque qui a voulu marquer le cheminement accompli dans l'évolution de la pensée sur l'évaluation en éducation dont notre association se veut le témoin sinon la mémoire, pour parler d'épistémologie de l'évaluation ?

*Comment l'aborder ?*

Mais alors, une nouvelle question se pose : pourquoi ne pas avoir invité un épistémologue des sciences pour faire le point sur cette discipline aujourd'hui ? C'était une possibilité que les organisateurs n'ont pas retenue. Nous n'aurons pas, par conséquent, un exposé sur une "théorie de la science" ni sur une "théorie de la connaissance" à l'aune desquelles nous aurions pu mesurer l'incongruité, l'insolence ou la vanité d'une formulation blasphématoire comme celle "d'épistémologie de l'évaluation". Nous n'entrerons pas, non plus, dans une réflexion systématique sur l'importance que l'on pourrait accorder aux modèles génétiques et constructivistes piagétiens ou aux conjectures réfutables de Popper pas plus que nous n'entreprendrons un essai d'appropriation de la conceptualisation épistémologique proposée par Bachelard autour de "la puissance de tremblement des concepts", par exemple. Mais nous ne nous interdirons pas de puiser dans ces références les éléments d'un questionnement épistémologique que nous revendiquons comme un droit et, plus encore, comme un devoir !

Nous situerons donc notre réflexion au niveau des préoccupations d'un "chercheur en évaluation" qui doute sans cesse de la scientificité des résultats que ses travaux produisent mais qui ne se satisfait pas de ce doute et qui cherche à en expliciter les éléments et surtout à les partager avec la communauté de chercheurs concernés.

Tout d'abord, il faudra bien commencer par situer une telle réflexion dans une perspective historique : il serait hasardeux et difficile de vouloir trouver le sens des productions actuelles sur l'évaluation sans les insérer dans une évolution. Certes, on ne va pas tenter de refaire l'impossible histoire d'une éventuelle succession d'idées sur l'évaluation en éducation (à laquelle plusieurs auteurs, surtout nord américains, se sont essayés avec des résultats mitigés sinon arbitraires).

En effet, on fait très rapidement le constat qu'il n'y a pas vraiment de "périodes" différenciées dans le domaine de l'évaluation, comme certains ont voulu nous le faire croire. On peut facilement vérifier autour de soi qu'aujourd'hui encore coexistent des pratiques d'évaluation qui relèvent de paradigmes ou de conceptions dont on parlait déjà en 1960 !

Il paraît plus intéressant d'aborder cette aventure en formulant une hypothèse générale qu'on tentera, bien sûr, d'étayer ou, au moins, d'illustrer. L'hypothèse est que le sens de l'évolution des travaux sur l'évaluation est à chercher dans le contexte de l'évolution des sciences de l'éducation et, plus généralement, des sciences humaines et sociales : une tendance à l'éclatement de ces disciplines expliquerait les difficultés actuelles d'ancrage des travaux qui en relèvent. Mais, en même temps, ce champ de recherche, de par son inscription sociale et institutionnelle forte, disposerait, aujourd'hui, de potentialités spécifiques de développement, à condition que la communauté scientifique qui s'y intéresse réfléchisse aux fondements qui justifient son existence. C'est autour de cette idée que va s'organiser la réflexion qui suit.

### Un exposé en trois temps

*Dans un premier temps*, on inscrira, rapidement, le propos dans l'évolution des conceptualisations les plus marquantes de l'évaluation. On rappellera les apports successifs de ce que l'on pourrait appeler les paradigmes fondateurs ainsi que le foisonnement des modèles qui ont théorisé l'évaluation et qui ont "produit", successivement, des courants de pensée organisés autour de l'évaluation en éducation ayant laissé, chacun, des "marqueurs" épistémologiques encore particulièrement prégnants dans ce champ de recherche.

*Dans un deuxième temps*, on proposera de situer la crise de la recherche sur l'évaluation dans un contexte d'éclatement des sciences humaines et sociales. On constatera, en effet, la contribution de l'évolution de ces sciences (dans le sens d'une dispersion des champs de recherche accompagnée d'une absence - ou d'une perte - de modèles unificateurs) à la déstructuration, voire à la confiscation du champ de l'évaluation.

*Dans un troisième temps*, on tentera de poser les termes d'un débat épistémologique sur l'évaluation. On verra que cette question se présente sous la forme de débats partiels et cloisonnés autour de la définition-même de la nature scientifique de l'évaluation, des choix épistémologiques souvent peu explicites qui sont réalisés, de l'hésitation à s'inscrire dans des paradigmes ouverts à la remise en question et à l'évolution, des conditions de la production de connaissances.

## I ère Partie - Inscrire le propos dans l'évolution des conceptualisations de l'évaluation

### En introduction : à la recherche d'un fil conducteur ?

Dans cette première partie, on cherchera, dans l'évolution des travaux sur l'évaluation, un fil conducteur allant de ce que nous appellerons les fondamentaux (en attendant de s'autoriser à parler de "paradigmes fondateurs") jusqu'au foisonnement des modèles qui ont théorisé l'évaluation.

Plutôt que de dresser un inventaire chronologique de plus, stipulant des périodes ou des phases historiquement datées - ce qui constitue souvent une entreprise artificielle (cf. Nadeau,) - on évoquera les principales conceptualisations qui ont "produit", successivement ou parfois parallèlement, des courants de pensée organisés et qui constituent des références incontournables pour toute réflexion sur l'évaluation. On évoquera :

- les fondamentaux (1) avec le rapport à la "mesure", l'épisode de la "docimologie", le courant de "l'évaluation des objectifs" ;
- les renouvellements (2) opérés par la prégnance du cognitivisme et l'apport des technologies informatiques ;
- une stabilisation provisoire des concepts (3) et la coexistence de grands "modèles" de l'évaluation.

## **I- Les fondamentaux**

### **I-1. L'évaluation et la mesure**

S'il est un concept fondateur et référent universel privilégié de l'évaluation, c'est bien celui de mesure. Les questions du genre "pourquoi mesurer", "quoi mesurer" et "comment mesurer" ont pu varier selon les époques et les courants mais le concept lui-même, "apprécier une dimension (longueur, surface, volume) en comparaison avec un étalon de même espèce", a constitué une des légitimités majeures des travaux liés à l'évaluation scolaire comme à l'évaluation de programme. La dimension quantitative de cette mesure s'est imposée, dans toute l'histoire de l'évaluation, comme la question majeure. Ainsi fallait-il distinguer la mesure de l'évaluation, cette dernière se réservant les fonctions d'appréciation qualitative, d'utilisation de la mesure pour prendre des décisions, apporter des améliorations ou des aides à l'apprentissage, mettre en place des boucles de régulation du fonctionnement des dispositifs ?

Pendant les années 1960-1980, des psychologues comme Campbell (1969, 1984) illustrent cette prévalence de la mesure en proposant une approche expérimentale des évaluations. Sur ce principe, deux sortes de pratique se sont installées dans le paysage de l'École et de l'orientation à travers la psychométrie (mesure de l'intelligence) et l'éduométrie (mesure des différents paramètres du processus d'apprentissage).

On les retrouve principalement dans deux approches apparemment voisines et pourtant bien différenciées :

**La psychométrie**, dans le domaine des Sciences Humaines et Sociales, remonterait, selon Parisot (1987), aux travaux de Galton (1879) qui l'avait définie comme "l'art d'imposer aux opérations de l'esprit mesure et nombre".

A partir du sens premier de "mesure de l'intelligence", la psychométrie est une notion qui a été attribuée à l'ensemble des pratiques d'évaluation ayant recours à la méthode des tests à partir de laquelle les individus sont classés, les uns par rapport aux autres, les tests habituels définissant la variance entre les individus comme la variance "vraie". La mesure effectuée détermine la position de l'individu étudié sur une échelle commune qui est censée représenter un trait commun à tous les sujets. Cardinet (1986) a fait une critique magistrale de cette approche, lorsqu'elle n'est pas complétée par d'autres formes d'évaluation, en montrant comment une telle mesure, extérieure à la situation d'apprentissage, ne pouvait pas prendre en compte l'ensemble des observations que seul l'enseignant peut récolter dans l'observation quotidienne de chaque élève.

**L'éduométrie** a été définie par la mesure des résultats de l'apprentissage au sens large que les Sciences de l'Éducation donnent à ce terme et s'est ainsi affranchie de la psychométrie. Pour les auteurs qui la décrivent (groupe suisse de la SSRE, autour de Cardinet et Bain, promoteurs en Europe de la théorie de la généralisabilité, initiée par Cronbach, 1972), elle évalue les progrès réalisés par un individu entre les étapes successives de son apprentissage et, par conséquent, les méthodes classiques de construction de tests ne s'y appliquent pas. Les recherches actuelles qui s'en réclament sont attachées à une plus grande "qualité de la mesure" et rejoignent certaines préoccupations de la docimologie.

### **I-2. La docimologie**

Piéron (1979) a défini la docimologie comme "l'étude systématique des examens" et de manière plus large ce champ de recherche s'est intéressé à tous les problèmes posés par la notation. Historiquement, cette orientation a vécu deux périodes distinctes : une phase négative (antérieure à la définition de Piéron) essentiellement préoccupée à critiquer expérimentalement les modes de notation et à dénoncer le manque de fidélité et de validité des examens ; une phase constructive, au cours de la quelle il a été proposé des méthodes de mesure plus objectives et plus rigoureuses, décrites par De Landsheere. On a pu penser que la docimologie constituait une science morte qui aurait dit ce qu'elle avait à dire et que son héritage étant établi, place était laissée à d'autres recherches sur les pratiques d'évaluation. On constate, néanmoins, avec l'exemple précédent de l'éducativité que cette approche continue à produire de la réflexion.

### **I-3. L'évaluation des objectifs**

Aucun bilan de l'apport des différentes écoles de pensée à la construction du concept d'évaluation en éducation ne peut ignorer le rôle important joué par la "pédagogie par objectifs", issue des travaux américains de Tyler et Bloom, à l'aube des années 60. Terme impropre, la pédagogie par objectifs fut, en réalité, la forme "pédagogisée" d'une théorie de l'action. Cet avènement de la "rationalisation de la pédagogie" a été marqué, on le sait, par les exigences de la productivité industrielle, exacerbées, à la fin de la deuxième guerre mondiale, par les nécessités de reconstruction rapide de l'Europe avec, notamment, le développement du Plan Marshall. Dans ce contexte, l'enseignement devait devenir "rentable" par la pédagogie par objectifs, comme la "direction par objectifs" devait rentabiliser la production. Dans un tel contexte idéologique, la fonction majeure de l'évaluation ne pouvait que devenir un contrôle de l'efficacité de la formation en référence aux objectifs que cette dernière affichait. Cette approche a réduit l'évaluation à devenir un instrument de comparaison entre des résultats prévus et des résultats obtenus, annulant ainsi tous les problèmes posés par la mesure de l'intelligence ou des capacités des sujets.

Mais cette approche s'est d'autant plus développée qu'elle a bénéficié du renfort inattendu des théories cognitives qui prévalaient à l'époque, en particulier celle du néo-béhaviorisme de Skinner pour lequel les connaissances, la personnalité et les comportements sont le résultat de ce que notre environnement nous fait vivre à travers un certain nombre de "renforcements". C'est à partir de cette thèse radicale que les renforçateurs sont devenus des éléments clés pour la modification des comportements. D'où l'idée particulièrement fructueuse de Bloom créant ses fameuses "taxonomies d'objectifs" qui ont servi de modèle et de référentiel à de nombreux protocoles d'évaluation des apprentissages.

Certes, aujourd'hui, la théorie cognitive explique différemment les processus d'apprentissage et la pédagogie par objectifs, ayant subi de nombreuses critiques, a perdu de son aura. Néanmoins, les signes de son ancrage dans les milieux de la formation n'ont pas disparu, en particulier dans l'organisation de l'entreprise. Ce qui explique la prégnance encore visible de l'évaluation des objectifs.

## **II- Les renouvellements**

### **II-1. La prégnance accrue du cognitivisme**

La synthèse des travaux sur l'évaluation en Europe et en Amérique du Nord, esquissée par Cardinet et Laveault (2001) met en valeur les avancées majeures de la réflexion sur l'évaluation des apprentissages suscitées par l'importance accordée aux apports des sciences cognitives, dans les années 90. Cette réflexion est illustrée par les auteurs à travers l'évocation du débat entre Skinner (déjà cité) et Flavell (1976) sur le fait de chercher à savoir si, pour étudier les effets des apprentissages, il convenait d'étudier "ce qui se passe entre les deux oreilles "(selon Flavell) ou si l'on pouvait "s'en passer", en observant seulement les comportements apparents (Skinner). En fait, il s'agissait là d'introduire le concept nouveau de métacognition qui allait bouleverser la façon de décrire la fonction de l'évaluation dans le développement du sujet. C'est dans le même sens que se sont également développées les idées sur l'autoévaluation ainsi que sur ce que certains formateurs et auteurs ont nommé "évaluation formatrice" pour spécifier le rôle et les stratégies du sujet apprenant dans l'utilisation des procédures d'évaluation à des fins de prise de conscience de ses propres processus d'apprentissage.

### **II-2. L'apport des technologies informatiques**

L'irruption des techniques d'information et de communication dans l'éducation a déjà été amplement analysée et fait aussi l'objet d'un bilan dans ce colloque. Les pratiques d'évaluation font partie des objets éducatifs les plus particulièrement touchés par ces technologies. En effet, les outils de stockage de l'information et de traitement automatique des réponses aux items, tant statistique que graphique, ont entraîné l'utilisation de questionnaires d'autoévaluation intégrés dans des sessions de formation à distance, donnant même l'impression à des enseignants et des formateurs qu'il ne leur était pas nécessaire de mobiliser de quelconques savoirs spécifiques pour concevoir des instruments d'évaluation. Le "savoir informatique" a souvent semblé pouvoir remplacer le "savoir évaluatif" : il était plus accessible, plus pratique et, de par sa logique binaire et sa force de calcul favorisant la quantification, il ne pouvait que renforcer ainsi l'illusion d'une plus grande scientificité. La recherche sur le sujet demande à être développée (on peut faire confiance aux équipes de Luxembourg pour cela) mais l'utilisation des TIC dans l'évaluation des enseignements par les étudiants dans les pays anglo-saxons et en Asie a donné lieu à des enquêtes et des travaux importants, notamment autour d'expériences réalisées dans des universités de Hong Kong (1999) : les analyses critiques de ces dispositifs ont souligné les problèmes posés à la maîtrise du processus d'évaluation par le maniement de tels outils (Theall, 2000). L'ère de l'illusion est, en tous cas, terminée. Mais la question qui reste posée et qui fait encore débat est la suivante : est-ce la technologie informatique qui influence la méthodologie d'évaluation ou est-ce le modèle de l'évaluation qui doit asservir la technologie informatique ?

### **III- Une stabilisation provisoire des concepts**

Si les années 60-80 ont vu naître ou se préciser des courants fondateurs d'une pensée organisée sur l'évaluation en éducation, on a assisté, dans les années 90, à un foisonnement "cohabitationnel" de conceptualisations de l'évaluation issues, sans aucun doute, des mouvements précédents. Ces conceptualisations ont favorisé la production d'un certain nombre de typologies fécondes pouvant donner à penser qu'on était arrivé à une sorte d'épanouissement de l'intégration de l'évaluation dans les problématiques d'apprentissage ou de pilotage des systèmes et dispositifs éducatifs.

#### ***Définitions stabilisées***

Il apparaît très clairement que, après le développement fulgurant de la notion d'"évaluation formative" lancée par Scriven et reprise par Bloom, les années 80 ont favorisé, tant dans les publications que dans l'offre de formation des enseignants, une diffusion très large de l'ensemble des fonctions pouvant être remplies par l'évaluation.

Ainsi, *l'évaluation diagnostique* (prédictive, de contexte) était développée par Stufflebeam (1980) et reprise par Cardinet (1986) ; *l'évaluation formative* était largement relayée par Allal, Cardinet et Perrenoud (1989) en Europe et par Scallon (1988), au Canada ; *l'évaluation sommative* donnait lieu à de nombreuses analyses, surtout dans ses distinctions avec l'évaluation formative ; plus récemment, *l'évaluation certificative*, décrite et problématisée particulièrement en France (cf. travaux de Dauvisis) ouvrait le champ de l'étude des liens entre l'école et la profession par la médiation de l'évaluation des aptitudes et, bientôt, des compétences. Des typologies équilibrées (en particulier autour du schéma "avant-pendant-après") ont été fréquemment présentées aux enseignants en formation, dans le but de montrer la complémentarité de ces fonctions.

#### **Oppositions banalisées**

Une autre dynamique a paru également intéresser les auteurs de cette période (Ardoino et Berger, 1989, Cardinet, déjà cité et De Ketele, et Bonniol, Vial) : celle des oppositions de paradigmes ou de concepts permettant de faire glisser des curseurs sur des axes dont les pôles sont opposés, comme, par exemple : contrôle / évaluation ; quantitatif / qualitatif ; résultat / processus ; externe / interne ; objet / sujet ; prescription / description... On reconnaît bien, dans ces différents couples opposés de concepts indiquant les choix alternatifs pouvant présider à la représentation et/ou à la pratique de l'évaluation, le contenu des débats d'idées qui ont eu lieu, pendant de longues années. Ces débats témoignaient notamment du dynamisme de la réflexion suscitée par l'évaluation dans le domaine éducatif et constituaient déjà une légitimation du statut d'objet de recherche spécifique de cette dernière.

### Synthèses publiées

En même temps, on a vu apparaître des "sommets" (bilans, classements, typologies) qui ont donné l'impression que la recherche sur l'évaluation était enfin capable de se regarder elle-même, avec recul et clairvoyance et qu'elle se trouvait à l'aube d'une affirmation encore plus large et stabilisée de son existence en tant que champ spécifique, en tant qu'objet isolable de la recherche en éducation. Quatre exemples, au moins, peuvent être évoqués du bien fondé de cette impression :

- l'ouvrage de Shadish, Cook et Leviton (1991) constitue une revue analytique de la littérature sur les théories de l'évaluation en général et sur les conceptualisations de l'évaluation de programme en particulier. Cet ouvrage de langue anglaise passe en revue les "théories" de l'évaluation publiées des années 60 aux années 90 aux Etats Unis. Le terme "théorie" est à comprendre, ici, au sens donné par les anglo-saxons, à la fois comme cadre général théorique et comme cadre de modélisation d'une pratique. Il s'agit d'une vision très vaste et particulièrement documentée de la connaissance produite par l'évaluation et de son inscription dans un contexte sociopolitique. La lecture de cet ouvrage nous libère des scrupules que nous pourrions avoir, les uns et les autres, lorsque nous entreprenons de conceptualiser et de théoriser des types de phénomènes, d'objets et de processus liés à l'évaluation : il y est constamment question des "théories" de l'évaluation et le panorama d'ensemble de ces "théories" décrit le niveau de scientificité atteint pour conduire la recherche sur l'évaluation.

- la thèse publiée de Rodrigues (1998-2002) contient également une somme bibliographique internationale importante à partir de laquelle l'auteur a esquissé l'existence de trois grands "paradigmes" capables de décrire et de classer les conceptions de l'évaluation en fonction de leurs "fondements axiologiques, éthiques, politiques et épistémologiques". On reviendra sur cette thèse par la suite.

- l'ouvrage de Bonniol et Vial (1997) opère un regroupement de textes "fondateurs" assortis d'analyses et de débats autour de trois "modèles" : "la mesure", la gestion" et "la problématique du sens", cette dernière catégorie n'étant pas encore renseignée et appelant de nouvelles applications.

- l'ouvrage que nous avons édité ("L'activité évaluative réinterrogée", 2001) constitue un bilan intermédiaire de la recherche sur l'évaluation durant les années 90, suscité par l'ADMEE Europe et s'est intéressé à montrer un certain nombre d'évolutions déjà perceptibles dans cette période. On a noté, en particulier, les tendances de développement des théories cognitivistes dans l'explication des liens entre apprentissage et évaluation ainsi que des modèles socioprofessionnels incitant de plus en plus à évaluer des "compétences".

\* A ces quatre sommets annoncés, on peut aujourd'hui en ajouter une cinquième : "Recherches sur l'évaluation en éducation : problématiques, méthodologies et épistémologie", paru à l'occasion des 20 ans de l'ADMEE...

#### *En conclusion : Elargissement du champ de l'évaluation*

Après l'enthousiasme des fondateurs des premières conceptualisations de l'évaluation (Scriven, Bloom, Cardinet,...) et l'âge d'or de la cohabitation des anciennes et nouvelles conceptualisations que nous venons d'évoquer, on s'est trouvé, à la fin des années 90, face à une recherche sur l'évaluation en éducation disposant

d'un ensemble de travaux de référence, de dispositifs institutionnels et professionnels suffisamment nourris et organisés pour que l'on puisse croire à la possibilité de construire un véritable thésaurus de l'évaluation, identifiable par un corpus diachronique et synchronique cohérent malgré sa diversité. On pouvait avoir l'impression, alors, qu'il existait une communauté scientifique adoptant un certain consensus sur les définitions, les questions, les méthodes utilisées pour interroger l'objet "évaluation". Mais en même temps, apparaissaient des apports nouveaux (la psychologie cognitive, les technologies informatiques, etc.) qui pouvaient constituer les prémices d'une "accumulation" portant le signe précurseur de l'avènement d'une crise scientifique, au sens de Kuhn (1972). En effet ce qui pouvait être ressenti comme un élargissement, en continuant à se développer, commençait à poser un nouveau problème : de l'élargissement à l'accumulation, de l'accumulation à l'éclatement, un nouveau processus était en route.

## **IIème Partie - Situer la crise scientifique de l'évaluation dans un contexte d'éclatement des sciences humaines et sociales**

**En introduction** : l'évaluation, miroir des sciences Sociales ?

Au cours des années 90, l'évaluation en éducation intéresse de plus en plus les chercheurs en sciences humaines et sociales. Elle est désignée comme analyseur important de l'évolution des systèmes éducatifs (évolution du niveau des élèves, efficacité des politiques et des dispositifs, questionnement des didactiques, rôle de l'école dans l'insertion professionnelle et sociale). De nouveaux travaux d'évaluation décomposent les problématiques traditionnelles en questions de plus en plus indépendantes (analyse des comportements évaluatifs, comparaisons internationales des performances des élèves, mise en relief des processus cognitifs liés à l'évaluation, enjeux sociaux, etc.). Enfin et surtout, les disciplines contribuant aux sciences de l'éducation produisent de plus en plus d'analyses de phénomènes appelées "évaluations" à partir de leurs propres modèles disciplinaires. La question se pose de plus en plus de savoir si le sens global qui pouvait être attribué à la notion d'évaluation n'est pas en train de se décomposer en autant d'opérations spécifiques et de recherches portant sur telle ou telle "découpe" d'un objet éducatif..

Mais si l'on regarde l'évolution actuelle des sciences humaines et sociales (S.H.S.) elles-mêmes, ne peut-on pas retrouver les mêmes questions ? Il paraît tout à fait intéressant et utile de réfléchir au statut épistémologique de l'évaluation dans un contexte plus large, celui d'une crise (au sens kuhnien évoqué plus haut) de ces sciences et d'examiner les formes et modalités des changements de paradigmes qui s'y déroulent. Dans ce vaste chantier de réflexion, on se limitera, ici, à proposer trois postulats :

- la "confiscation" du champ de l'évaluation opérée par certaines SHS
- le contexte général "d'éclatement" des sciences humaines et sociales elles-mêmes
- le repositionnement perceptible de la recherche spécifique sur l'évaluation

## **1- Le champ de l'évaluation "confisqué" par les sciences humaines et sociales ?**

L'apport des recherches de différentes disciplines de S.H.S. à l'étude des phénomènes liés à l'évaluation depuis plus d'une décennie est remarquable. Les deux exemples les plus marquants sont certainement la sociologie et la psychologie sociale, mais on pourrait aussi citer l'économie, les sciences politiques, les sciences de la communication et, évidemment, la psychologie cognitive, déjà abordée précédemment.

### **1-1. Contributions de la sociologie**

On se limitera à trois exemples de traitement d'évaluation par la sociologie.

#### ***- explication sociale de l'évaluation scolaire***

Selon Merle (1998), les recherches menées sur l'évaluation scolaire par les sociologues peuvent se classer autour de trois grandes orientations :

. les théories de la reproduction (Bourdieu, Passeron) qui ont exercé une influence dominante durant les années 70, répondent à la question du fondement des inégalités sociales de réussite scolaire en stipulant que l'école est un lieu de transformation des inégalités culturelles en inégalités scolaires. L'évaluation en classe ainsi que l'examen ne sont, dans cette perspective, analysés que comme des "illusions scolaires à la légitimation des élites sociales" (Bourdieu, 1979).

. à partir des années 80, plusieurs orientations semblent coexister. La première se préoccupe des relations maître-élève et, tout particulièrement, des modalités concrètes de fabrication des notes. La seconde orientation de recherche, essentiellement statistique combine une perspective sociodémographique (intéressée au mode de sélection des élèves) et une approche socioéconomique (intéressée par la question de l'efficacité). Cette orientation a donné lieu à de nombreux travaux sur l'effet du contexte de scolarisation (Duru-Bellat, Mingat, 1997).

. une troisième approche qui complète les deux autres consiste à s'intéresser aux "variables de position". Il s'agit, pour les auteurs concernés, d'expliquer les résultats des élèves par le produit d'histoires personnelles très tôt différenciées et par la diversité des trajectoires scolaires qui en découlent (Lahire, 1995).

#### ***- une application à l'étude de l'efficacité de l'école***

Les travaux sur "l'école efficace" ont marqué les années 70 et 80 en Europe et dans les pays de l'OCDE : on peut avoir une idée de l'étendue des recherches suscitées par ce courant en consultant les rapports de Scheerens (1989 ; 1995) ainsi que les travaux de Grisay (1993). Ils ont influencé de manière importante les conceptions de l'évaluation des dispositifs et des institutions en généralisant et en vulgarisant le recours aux "indicateurs" qui sont devenus aujourd'hui couramment utilisés.

#### ***- étude de l'évaluation comme pratique sociale***

Trois approches distinctes peuvent être repérées :

. l'analyse des actes et des procès d'évaluation qui distingue l'évaluation des agents et l'évaluation des actions et les situe dans leurs contextes sociaux (Barbier, 1985)

. l'analyse de l'efficacité de l'évaluation et de sa valeur épistémologique qui subordonne la qualité de cette évaluation à l'existence, même implicite, d'une théorie sociologique qui "informe une représentation du social" (Demailly, 2001)

. l'analyse des fonctions et des effets de l'évaluation dans la "fabrication de l'excellence scolaire" (Perrenoud, 1995).

On voit bien, à travers ces différents regards, comment l'approche sociologique choisit d'éclairer directement et unilatéralement certaines variables contextuelles de l'acte d'évaluer, au risque de confisquer le sens du processus évaluatif au profit du seul paradigme sociologique.

### **1-2. Contributions de la psychologie sociale**

La psychologie sociale a également revendiqué la construction de savoirs sur l'évaluation (Monteil, 1989). En particulier, des recherches sur les pratiques de l'évaluation ont mobilisé les théories de l'attribution (étudiant la manière d'attribuer une cause à un comportement observé) et ont pu ainsi dégager les biais des jugements évaluatifs. Il a été montré, également, la tendance à surestimer les facteurs personnels au détriment du contexte social. Dans cet esprit, les chercheurs ont même tenté de démontrer l'existence d'une norme sociale, la "norme d'internalité" qui influence les jugements évaluatifs (Dubois, 2001) : ceux-ci favoriseraient les comportements "internes" (des individus qui auraient tendance à considérer que les résultats qu'ils obtiennent dans les évaluations découlent directement de ce qu'ils sont ou de ce qu'ils ont fait) par rapport aux comportements "externes" (des individus qui pensent dépendre de forces extérieures et indépendantes de ce qu'ils sont ou de ce qu'ils font). Les évaluateurs, en général, jugeraient plus favorablement les sujets internes que les sujets externes et, fait aggravant, les sujets externes se retrouveraient plus fréquemment dans les milieux défavorisés que les sujets internes. Cette norme sociale, si elle est avérée, bouleverse quelque peu le paysage de l'évaluation et de l'éducation en général, dans la mesure où elle démontrerait que les grandes réformes pédagogiques depuis un demi-siècle tendant à valoriser la responsabilité et l'autonomie de l'élève seraient responsables d'injustices et d'échecs.

L'exemple de ces deux disciplines, la sociologie et la psychologie sociale, (exemple qui aurait pu être renouvelé, comme cela a été dit, dans les autres disciplines des SHS) illustre une tendance forte à traiter les problématiques d'évaluation de manière isolée et ponctuelle, sous couvert de traiter l'incidence de telle ou telle variable sur les phénomènes liés à celle-ci. Chaque étude se réclame de la légitimité de son inscription dans un paradigme disciplinaire reconnu sans relier son questionnement à la complexité inhérente à toute opération d'évaluation faisant interagir des partenaires différents (= évaluateurs et évalués, individus et collectifs) dans des situations régies par des processus de communication, de conflits et de résolution de problèmes. La notion de "situation globale d'évaluation" disparaît ici et les évaluateurs responsables de la conduite d'une procédure de régulation de l'action évaluée font place aux chercheurs experts dans une "discipline-outil".

Est-ce à dire, à ce moment de la réflexion, qu'il faut prédire la fin de ce qu'on va appeler, dans un instant, les "paradigmes de l'évaluation" et l'avènement des "paradigmes disciplinaires" appliqués à l'évaluation ?

## **2- En filigrane : le contexte général "d'éclatement" des sciences humaines et sociales elles-mêmes**

Pour répondre à la question on peut d'abord se demander si le morcellement des problématiques d'évaluation évoqué précédemment ne constitue pas déjà un reflet d'un problème plus général qui concerne l'évolution des S.H.S. elles-mêmes ?

### **2-1. Multiplication des champs de recherche**

Dans un récent numéro spécial de la Revue Sciences Humaines (janv. 2006), Dortier soutient que "les sciences humaines sont entrées dans une nouvelle période où la multiplication des champs de recherche et des savoirs locaux, en même temps que l'absence de modèles unificateurs rendent paradoxalement impossible toute vision globale... L'esprit de système semble peu à peu abandonné partout au profit des études locales, plus modestes". Le bilan ainsi introduit s'appuie sur des analyses de différentes disciplines concernées par cette évolution comme les sciences cognitives qui étendent progressivement leurs modèles à des disciplines voisines (sociologie, économie, éthologie, thérapie... "cognitives"), la sociologie de plus en plus fascinée par "l'individu" (étudié à travers le déclin des normes et le renforcement de la sphère privée), la linguistique qui s'est subdivisée en une mosaïque de courants (la sémantique, la linguistique textuelle des discours ou celle de l'énonciation...). Le diagnostic qui clôt cette analyse consiste à postuler que, si nombre de reconfigurations en cours restent confinées dans des bassins de connaissances spécialisés, c'est parce qu'il manque une impulsion aux synthèses, bilans, cartographies des savoirs, mais aussi et surtout aux outils conceptuels destinés à réarticuler entre eux les savoirs locaux.

### **2-2. Un repositionnement de la recherche spécifique sur l'évaluation ?**

C'est une question à prendre au sérieux, dans le contexte de restructuration des disciplines universitaires autour des nouveaux masters européens. Et cette question en appelle d'autres.

L'évaluation en éducation fait, aujourd'hui, l'objet d'une demande très forte, à la fois institutionnelle (avec les évaluations commanditées par les ministères et les organismes de veille nationaux et internationaux) et professionnelle (avec les attentes des entreprises et des organisations vis-à-vis des dispositifs de certification, d'assurance-qualité, d'audit et de contrôle de leur fonctionnement). Paradoxalement, un tel champ d'investigation qui avait mis, comme on l'a vu, un demi-siècle pour construire un début de dialogue entre conceptualisations complémentaires et devenir un objet de recherche distinct, au sein des Sciences de l'Éducation, serait-il en situation d'éclatement, en partie sous la pression des disciplines de S.H.S. en quête de nouveaux sujets porteurs ?

### **2-3. Le rôle des laboratoires de sciences de l'éducation**

En effet, dans un tel contexte de demande sociale et professionnelle forte, la tentation est grande pour toute équipe universitaire dotée d'une spécificité disciplinaire ayant recours aux outils sociologiques, économiques ou psychologiques, de se proposer, en concurrence avec les sciences de l'éducation, pour l'évaluation d'une politique, d'un dispositif, voire d'une méthode d'éducation ou de formation. Il s'agit là d'un enjeu important pour les sciences de l'éducation, aujourd'hui : ne sont-elles pas assez puissantes pour conserver en leur sein des équipes de recherche (ou des laboratoires) dédiées à l'évaluation ? Ces équipes ne sont-elles pas susceptibles d'attirer à elles des collaborations disciplinaires complémentaires autour d'une problématique complexe et intrinsèquement liée aux processus d'enseignement-apprentissage et de régulation des systèmes éducatifs, comme l'est l'évaluation ? On perçoit ici que le problème, s'il existe, ne provient pas du champ-même de l'évaluation et de son développement comme objet d'étude mais des Sciences de l'Education et de leur positionnement dans l'ensemble des disciplines de S.H.S. A moins que les Sciences de l'Education, oubliant leur pluralité, par mimétisme des autres sciences humaines et sociales, ne jouent elles-mêmes ce même jeu de dispersion ?

## **IIIème Partie - Poser les termes d'un débat épistémologique sur l'évaluation**

### **En introduction : les questions scientifiques qui préoccupent les chercheurs en évaluation**

Dans une telle situation, peut-on repérer des questionnements qui traversent la recherche sur l'évaluation ? et ont-ils encore un sens ? Je voudrais poser ici l'idée qu'il s'agit là de notre responsabilité, je veux dire ce celle de notre communauté scientifique rassemblée au sein de l'ADMEE qui consiste à relancer un débat épistémologique qui semble manquer à certains d'entre nous. Les points communs de nos préoccupations de chercheurs en évaluation ne manquent pas, en effet. J'en choisirai quelques uns autour desquels je souhaiterais, dans cette occasion qui m'est donnée, pointer quelques éléments du débat qui pourrait être le nôtre :

- caractériser la nature scientifique de l'évaluation et se référer à une théorie
- définir son objet entre science et discipline
- identifier et assumer les choix épistémologiques directement ou indirectement opérés (démarquer son objet du "sens commun", se situer entre "causalisme" et "finalisme" de l'évaluation, s'inscrire clairement dans tel ou tel "paradigme"
- s'interroger sur la nature des connaissances produites

### **3-1. Caractériser la nature scientifique de l'évaluation, entre activité appliquée et besoin de théorie**

Ce genre de question peut difficilement se poser en dehors des sciences de l'éducation en général. Or on entend dire, en effet, que ces dernières sont toujours dans leur phase

"critique", "préscientifique" et qu'elles n'ont pas encore atteint le stade de la "science normale", c'est-à-dire celui des lois et des généralisations. En fait, il s'agit de savoir si, en quoi et à quelles conditions elles peuvent sortir de l'emprise hégémonique de l'approche prescriptive que constitue le modèle scolaire ou "pédagogique" pour accéder à un statut objectif distinct de son objet d'étude.

Mais, au regard du problème que se posent les sciences de l'éducation, la recherche en évaluation recueille des avis partagés. Pour certains auteurs (cf. Wortman, 1983), elle apparaît encore plus nettement comme une activité appliquée, largement (et peut-être malheureusement) a-théorique.

D'autres, cités dans l'ouvrage de Shadish et al. (1991) sur l'évaluation de programme, déplorent que de nombreux ouvrages sur l'évaluation soient des listes a-théoriques de méthodes. "Les théories de l'évaluation sont nécessaires", selon eux, pour "construire une connaissance de la valeur des programmes". Une telle théorie (de l'évaluation de programme) doit "explicitement décrire et justifier les cinq composantes suivantes reliées entre elles de manière logique : la programmation sociale, la construction de connaissance, l'attribution de valeur, l'utilisation de la connaissance et la pratique d'évaluation".

Dans ce cadre théorique, Shadish et al. étudient, à partir d'une revue de la littérature nord américaine de la deuxième moitié du siècle dernier, les caractéristiques de la construction de connaissances par (ou sur) l'évaluation. D'après les auteurs, les connaissances élaborées par les évaluateurs et qui émergent de leur analyse ont pour base :

- . "une ontologie, étude de la nature ultime de la réalité observée" (par exemple, les caractéristiques des apprentissages évalués)

- . "une épistémologie, étude de la nature, des origines et des limites de la connaissance élaborée" (par exemple, les causalités, les corrélations, les généralisations)

- . "une méthodologie", étude des techniques pour la construction de la connaissance (en l'occurrence, le recours aux méthodes des sciences sociales).

On peut évidemment remettre en question l'attribution du statut de "théorie" à un tel modèle qui pourrait aussi être interprété comme la seule schématisation d'une démarche et des consignes qui s'y rattachent.

Les préoccupations théoriques ne sont pas absentes, non plus, dans les écrits actuels. Par exemple, dans le recueil qui vient de marquer les 20 ans de l'ADMEE (Figari, Mottier-lopez, 2006), les références aux théories de la mesure (la généralisabilité, la théorie classique des tests, la théorie de réponses aux items invoquées par J-G Blais) figurent en bonne place. Nicole Dubois évoque, dans sa contribution, le recours à des théories implicites intervenant dans l'activité d'inférence qui constitue, selon elle, le cœur de la démarche évaluative. En tous cas, un débat sur le statut "théorique" de l'abstraction et de la conceptualisation préalable à la construction de protocoles d'évaluation reste d'actualité au sein de l'ADMEE.

### **3-2. Définir son objet entre science et discipline**

*Par rapport à la notion de "science", d'abord. Nous sommes dans l'ensemble persuadés que les sciences humaines et sociales semblent bien dans l'incapacité de saisir l'enchaînement des phénomènes sociaux à travers des lois régulières,*

exprimables sous une forme mathématique. Mais il est vrai aussi que certains chercheurs se raccrochent à l'idée qu'il reste la possibilité de mettre en évidence des régularités statistiques (qui donnent l'impression d'une compréhension scientifique). C'est à peu près dans les mêmes termes que se pose le problème lorsque l'on veut définir la scientificité des travaux d'évaluation. On ne peut, ici, que pointer cet autre objet de débat, permanent et inévitable, qui consiste à opposer l'évaluation-jugement, soumise au subjectivisme sans partage, à une évaluation-observation, à la recherche d'objectivité et de rigueur méthodologique.

*Par rapport à la notion de "discipline"*, ensuite. Cette question qui se pose d'abord pour les sciences de l'éducation a été magistralement traitée par nos collègues suisses Rita Hofstetter et Bernard Schneuwly (2001). Les auteurs emploient le terme de "champ disciplinaire" pour définir les sciences de l'éducation et décrivent un processus de "disciplinarisation" de celles-ci à partir d'un "champ professionnel" par opposition aux processus de disciplinarisation en rapport avec des savoirs uniquement académiques constitutifs d'autres disciplines. On pourrait très bien appliquer à l'évaluation la définition d'un processus d'élaboration de savoirs autour de champs professionnels préalablement constitués et en réponse à des demandes sociales d'ordre socio-professionnel et politico-administratif".

De là à oser appeler "discipline" l'évaluation, il y a certes un grand pas que nous n'oserions pas franchir ici, faute de preuves, et qui pourtant semble franchi et dépassé, en Amérique du Nord à propos de "l'évaluation de programme". Le récent et très sérieux article de Nathalie Dubois et Richard Marceau (2005) paru dans la Revue Canadienne d'évaluation de programme se propose d'étudier l'état actuel de la "*discipline* Evaluation de Programme", à partir d'un modèle d'analyse s'inspirant de Kuhn et de Lakatos. Il s'agit, pour les auteurs, de dresser "un état des lieux théoriques de l'évaluation comme discipline à la remorque d'une révolution scientifique qui n'en finit pas". La lecture de ce article, particulièrement bien documenté, mériterait elle-aussi, discussion : je la verse au dossier que je propose que nous ouvrons, à l'ADMEE.

### **3-2. Des questions sur les "choix épistémologiques"**

#### **- objet de débat n°1 : démarcation entre objet scientifique et "sens commun"**

Une des difficultés qui se présentent au chercheur en évaluation est celle qui consiste à opérer nettement cette distinction, lorsqu'il s'agit d'étudier le comportement évaluatif intégré aux activités scolaires ou formatives ancrées dans une relation "maître-élève" quotidienne et apparaissant comme attitude naturelle "allant de soi". Y a-t-il lieu d'opposer savoir scientifique et sens commun (comme le préconise Popper) ou comme le définit Bachelard à travers la notion de "coupure épistémologique" ? ou bien faut-il suivre le modèle phénoménologique qui considère le sens commun comme une première conceptualisation de la vie sociale ? Ce type de choix épistémologique n'est pas fréquemment exprimé dans la réflexion évaluative : il relève plutôt de l'implicite tout en contribuant à la perception "floue" qu'on peut avoir de la méthode évaluative.

**- objet de débat n° 2 : hésitation entre causalisme et finalisme**

Le chercheur en évaluation manipule en permanence la notion de "cause" à travers l'étude du lien "cause/effet". Il est alors confronté à un autre problème épistémologique important traité, notamment, par Collin (2000) : celui du débat entre causalisme et finalisme.

Il peut en effet, comme Durkheim l'a posé, affirmer l'existence d'une "causalité sociale" de la même manière qu'on peut affirmer qu'il y a une "causalité naturelle". Cela veut dire que "cette causalité sociale détermine des comportements individuels, y compris les représentations que les individus se font de leurs propres comportements". Mais on peut ne pas vouloir ignorer les finalités conscientes de l'action humaine, particulièrement prégnantes dans le processus évaluatif. C'est l'idée que les choses existent en vue d'un but qui nous concerne : dès que l'évaluateur ou l'évalué doit déterminer ce que "doit être son action", il réintroduit, nécessairement, une téléologie. On connaît la critique du finalisme faite par Spinoza dans sa phrase célèbre : "nous ne désirons pas une chose parce que nous la trouvons bonne mais nous la trouvons bonne parce que nous la désirons". Il n'empêche que le problème est quasi insoluble, à moins qu'on ne se satisfasse de la synthèse que fait Weber des deux approches : "comprendre les comportements humains, c'est les comprendre comme des moyens rationnels (donc observables) déployés en vue d'atteindre une certaine fin.

**- objet de débat n° 3 : recours aux "paradigmes de l'évaluation"**

L'expression "paradigmes de l'évaluation" a été employée par De Ketele (1993, 2001) pour référer les différentes pratiques à des courants de pensée qui leur confèrent un sens privilégié. L'auteur en avait recensé 10 (parmi lesquels par exemple, le paradigme docimologique, le paradigme de l'évaluation au service de la décision, le paradigme de l'évaluation répondante, etc.) auxquels il a suggéré d'ajouter (2001) celui, plus englobant et selon lui, en cours de construction, de "l'évaluation-régulation centrée sur l'action située". Le choix et la formulation de ces paradigmes reflètent de manière éclatante l'état d'esprit des années 80 présentées précédemment comme une période de stabilisation et de cohabitation de concepts d'origines diverses mais contribuant à décrire de manière de plus en plus complexe les processus de l'évaluation abordés en tant que tels. On peut douter du consensus qui pourrait exister aujourd'hui sur l'acceptation de certains de ces référents transversaux aux champs de savoirs disciplinaires alors que les paradigmes "sociologique" et "économique", également cités par De Ketele, restent, quant à eux, de plus en plus invoqués.

Mais si on adhère à la définition que donne Kuhn (1983) du paradigme comme étant "ce que les membres d'une communauté scientifique possèdent en commun" et à son corollaire qui est "qu'une communauté scientifique se compose de chercheurs qui se réfèrent au même paradigme", on a plus de difficulté à parler de "paradigmes de l'évaluation" du fait de l'éclatement des modèles et des concepts tel qu'évoqué précédemment. C'est pourtant la démarche qu'entreprend Pedro Rodrigues dans sa thèse encyclopédique dont il a donné des aperçus dans les publications de l'ADMEE, y compris dans celle des 20 ans. L'auteur a, dans le sens contraire des modes actuelles, classé de très nombreuses publications, particulièrement anglo-saxonnes, autour de trois "logiques d'évaluation en relation avec des conceptions élargies du monde et de la connaissance" (2006). Rodrigues distingue "le paradigme objectiviste-positiviste (observation externe et comparaison), le paradigme subjectiviste-

phénoménologique (évaluation par les acteurs eux-mêmes) et le paradigme dialectique-interactif" (coévaluation). Dans le même esprit, Cardinet (1987) avait utilisé une typologie très proche de celle de Rodrigues en proposant de distinguer entre "évaluation interne, externe et négociée".

A l'image des sciences humaines et sociales pour lesquelles Dortier se demande si de nouveaux paradigmes ne sont pas en train de prendre le relais des anciens, même s'ils n'ont pas été portés par de "grandes figures" ou des œuvres marquantes (comme ce serait le cas du paradigme cognitif), on est tenté de suggérer, comme De Ketele, que l'évaluation peut aussi espérer en contenir ou en produire. Le concept de "crise" utilisé par Kuhn peut conforter cette idée : en effet les crises concernent les périodes où les scientifiques sont confrontés à l'incapacité de leur modèle (ou paradigme) à résoudre une énigme. Si la crise gagne de l'importance et surtout perdure, la validité du modèle est remise en cause par l'émergence d'un nouveau paradigme qui représente une réponse possible à la crise. Parmi ces nouveaux paradigmes qui pourraient émerger des travaux de ces dernières années, celui formulé par De Ketele de "l'évaluation-régulation centrée sur l'action située" ne demande qu'à être collectivement examiné. Il semble, en effet, en relation avec la problématique naissante d'une "évaluation située", intégrée à des pratiques réflexives, au sens que donne Jorro (2005) à ce concept inscrit dans "la pensée critique- régulatrice".

L'évaluation est alors appréhendée dans un champ de significations, d'activités et de ressources socialement organisées, dans et par lequel elle se développe. Réciproquement, l'évaluation contribue au développement de "communautés de pratique" (Mottier Lopez, 2006) notamment par la mise en tension entre les valeurs, perceptions et positionnements individuels, et les normes et pratiques des différentes communautés convoquées.

On voit encore, ici, que la question des paradigmes n'est pas close, non plus.

### **3-4. Des questions sur la production de connaissances**

Les débats sur la scientificité de l'évaluation ont toujours confondu deux démarches pourtant distinctes :

- 1- la recherche sur les phénomènes induits par la pratique évaluative dans les situations d'enseignement-apprentissage ou de fonctionnement de l'école ;
- 2- l'évaluation elle-même comme appréciation d'une performance, des résultats d'une méthode ou d'un dispositif.

La première activité relève d'une conceptualisation préalable à l'élucidation de problèmes tels que l'influence des contextes, le rôle des acteurs, l'incidence des méthodes. Elle réalise des investigations de portée générale, transposables, généralisables selon les règles fixées par les méthodes de sciences humaines et sociales et propose des résultats critiquables par la communauté des chercheurs.

La seconde constitue une pratique de régulation intégrée au pilotage des dispositifs éducatifs : la connaissance qui en est issue est locale, unique, mouvante et non reproductible. Ses résultats sont indicatifs, provisoires et utilisables principalement par les acteurs (élèves/apprenants, enseignants, parents, décideurs).

Le lien entre les deux activités consiste à utiliser la conceptualisation de la première au profit des choix méthodologiques de la seconde : là est peut-être la principale source de confusion.

Mais si cette distinction est opérée, la question épistémologique est déjà considérablement simplifiée : la première activité s'inscrit dans une démarche scientifique alors que la seconde reste une pratique (pouvant être à son tour étudiée par la première).

L'activité de recherche sur les phénomènes induits par la pratique évaluative a vocation à produire des connaissances générales à propos de l'influence de certains facteurs sur le développement, la réussite d'un type d'apprentissage, de dispositif ou de politique (par exemple "l'effet-maître" ou "l'effet établissement" sur la réussite des élèves). Elle tentera également de faire progresser la connaissance des phénomènes liés à la pratique évaluative (par exemple l'influence sociale qui détermine les jugements évaluatifs ou le rôle de l'évaluation métacognitive dans le renforcement des apprentissages). Ces savoirs d'un type particulier, parce qu'induits très souvent des pratiques professionnelles qui en avaient besoin, une fois constitués, donnent lieu à conceptualisation, à théorisation, à modélisation et son soumis, comme d'autres types de savoirs en cours de constitution, à toutes les vérifications et les "falsifications" souhaitables. Ils sont nommés "savoirs d'action" (Jorro, 2006), ils relèvent d'une "science de l'action" (Duru-Bellat), il sont "d'ordre métacognitif" (Noel) et revêtent des formes conceptuelles pragmatiques construites pour l'action" (Mayen) ce qui rejoint l'expression de "savoirs d'utilité" de Freeman.

Cette production quasi "cyclique" de savoirs donne à penser que la recherche sur l'évaluation, si elle ne peut prétendre être nomothétique, comporte, en revanche, une dimension heuristique renouvelable en permanence : chaque résultat, en modifiant le milieu (la situation) qui l'a produit, provoque un nouveau questionnement.

## **CONCLUSION**

### *D'une épistémologie en débat à la recherche de nouveaux paradigmes ?*

On a pu voir la place de la recherche sur l'évaluation en éducation dans le contexte général de "pensée éclatée" qui semble être vécu dans les sciences humaines et sociales, contribuant à une certaine diversification des approches scientifiques de l'évaluation.

On a noté la tendance de certaines disciplines à traiter le sujet de l'évaluation à partir de leurs paradigmes internes et les conséquences possibles sur le désengagement des laboratoires vis-à-vis de ce sujet.

Mais, contrairement à cette analyse, une autre analyse, issue d'une revue de travaux nord américains, notamment, conclut à l'existence d'un espace (attesté, insuffisant, nécessaire) pour théoriser l'évaluation. Il existerait une voie pour construire une "théorie pragmatique" de l'évaluation en partie tributaire et en partie indépendante des débats scientifiques qui animent aujourd'hui la communauté des chercheurs en sciences humaines et sociales. Il appartiendrait aux chercheurs spécialisés sur ce sujet de rassembler les pièces (déjà nombreuses) du puzzle en un tout cohérent, de recoller les morceaux d'une "évaluation en miettes" (Ardoino, Berger, 1989) ou, pour généraliser, d'un "savoir en miettes".

On a soulevé un problème qui semble préoccupant et qui est celui du combat entre deux paradigmes majeurs de la recherche sur l'ensemble des processus de l'évaluation : *le paradigme "disciplinaire"* qui jouerait un rôle "disjonctif" comme le dirait Lerbet (1995) en découpant l'objet "évaluation" en autant de facettes que de disciplines intéressées à étudier telle ou telle de ses variables et *un paradigme "interdisciplinaire"* qui aurait la vocation de le traiter dans sa signification globale, complexe, interactive, située, ce que Lerbet qualifierait d'approche "conjonctive" de la démarche scientifique.

Ce problème invite à réfléchir, encore, toujours et plus que jamais, sur le statut épistémologique de l'évaluation. A la dispersion des champs de recherche ne peut-on pas opposer la mise en évidence de principes unificateurs, ancrés dans la théorie des sciences sociales et adaptés à la recherche de nouvelles connaissances sur les processus d'évaluation remis en question par les avancées des recherches récentes ou naissantes sur la compétence et l'expérience, par exemple ? Oui, décidément, nous plaidons pour une refondation épistémologique concertée de la recherche sur l'évaluation.

**Bibliographie**

- ADMEE Europe, Actes (sur CDROM) du XVIIème Colloque sur le thème "L'évaluation des compétences entre reconnaissance et validation des acquis de l'expérience", Université de Lisbonne, 18-20 nov. 2004.
- Allal, L., Cardinet, J., Perrenoud, P. (1989). *L'évaluation formative pour un enseignement différencié*. Berne : Peter Lang.
- Ardoino, J., Berger, G. (1989). *D'une évaluation en miettes à une évaluation en actes*. Matrice, ANDSHA.
- Barbier, J.-M. (1985). *L'évaluation en formation*. Paris : PUF.
- Becker, G.S. (1964). *Human Capital. A theoretical and empirical analysis with reference to education*. N. York : Columbia University Press.
- Bonniol, J.-J., Vial, M. (1997). *Les modèles de l'évaluation*. Bruxelles : De Boeck.
- Bourdieu, P. (1979). *La noblesse d'état*. Paris : Minuit.
- Campbell, D.T. (1969). *Reforms as experiments*. American Psychologist, 24.
- Campbell, D.T. (1984). Can we be scientific in applied social science ? in Evaluation studies annual review. Beverly Hills : Sage.
- Cardinet, J. (1986). *Evaluation scolaire et mesure*. Bruxelles : De Boeck Université.
- Cardinet, J. (1986). *Pour apprécier le travail des élèves*. Bruxelles : De Boeck Université.
- Cardinet, J. (1990). Evaluation externe, interne ou négociée ? *Hommage à Jean Cardinet*. Fribourg : Delval.
- Cardinet, J., Laveault, D. (2001). L'activité évaluative en éducation: évolutions des préoccupations des deux côtés de l'atlantique. In Figari, G., Achouche, M. (Eds.). *L'activité évaluative réinterrogée*. Bruxelles : De Boeck.
- Collin, D. (2000). *La théorie de la connaissance*. Paris : l'Harmattan.
- Cronbach, L.J. et al. (1972). *Theory of generalisability for scores and profiles*. New York : John Wiley.
- Dauvisis, M.-C. (1999). Valeurs latentes des pratiques d'évaluation certificative. *Spirale*, 23, 231-241.
- De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue Française de Pédagogie*, 113, 59-80.
- De Ketele, J.-M. (2001). Evaluation des compétences : une problématique émergente. In Figari, G., Achouche, M. (Eds.). *L'activité évaluative réinterrogée*. Bruxelles : De Boeck.
- Demailly, L. (2001). Evaluer les évaluations. In Figari, G., Achouche, M. (Eds.) (2001). *L'activité évaluative réinterrogée*. Bruxelles : De Boeck.
- Dortier, J.-F. (2006). "L'intelligence dispersée". *Sciences Humaines*, 167, numéro spécial "La pensée éclatée", 22-23.
- Dortier, J.F. (2004). *Le dictionnaire des sciences humaines*. Paris : Sc. Humaines Editions.
- Dubois, N. (2001). Internalité et évaluation scolaire. In Figari, G., Achouche, M. *L'activité évaluative réinterrogée*. Bruxelles : De Boeck Université.
- Dubois, Nath., Marceau, R. (2005). Un état des lieux théorique de l'évaluation : une discipline à la remorque d'une définition scientifique qui n'en finit pas. *The Canadian journal of Program Evaluation*, 20, 1 (1-36).
- Duru-Bellat, M., Mingat, A. (1997). La gestion de l'hétérogénéité des publics d'élèves au collège. Cahiers de l'IREDU, 59.
- Figari, G. (1996). *Evaluer : quel référentiel ?* Bruxelles : De Boeck Université.
- Figari, G. (2001). Us et abus de la notion de référentiel. In Figari, G., Achouche, M. (Eds.). *L'activité évaluative réinterrogée*. Bruxelles : De Boeck Université.
- Figari, G., Achouche, M. (2001). *L'activité évaluative réinterrogée*. Bruxelles : De Boeck Université.

Figari, G., Rodrigues, P., Alves, P., Valois, P. (Eds), (2006). *Evaluation des compétences et apprentissages expérimentiels (Savoirs, modèles et méthodes)*. Lisbonne : Educa.

Figari, G., Mottier-Lopez, L, (eds) (2006). *Recherche sur l'évaluation en éducation*. Paris : L'Harmattan.

Flavell, J.-H. (1976). Metacognitive aspects of problem solving. In L.B. Resnick (Ed.). *The nature of intelligence*. Hillsdale, Erlbaum,

Grisay, A. (1993). *Le fonctionnement des collèges et ses effets sur les élèves de 6<sup>e</sup> et 5<sup>e</sup>*, *Dossiers Éducation et Formations*, 32.

Ha, T.S., Maarsh, J. (1999). Implementation of a system for On-line evaluation of teaching. In Balasubramanian (ed.). *Teaching and learning quality management through use of information technology*. Hong Kong, University Grants Committee.

Hofstetter, R., Schneuwly, B. (2001). *Les sciences de l'Éducation en Suisse, Evolutions et prospectives*. Berne : CEST.

Jorro, A. (2005). Réflexivité et auto-évaluation dans les pratiques enseignantes. *Mesure et évaluation en éducation* ; 27 ; 2 ; 33-47.

Jorro, A. (2006) Devenir ami critique. Avec quelles compétences ? Quels gestes professionnels ? *Mesure et évaluation en éducation*, n°29. (31-44). Québec.

Kuhn, T. (1972, 1983). *La structure des révolutions scientifiques*. Paris : Flammarion.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In *Criticism and the growth of knowledge*. Cambridge University Press.

Lahire, B. (1995). *Tableaux de famille. Heurs et malheurs scolaires en milieux populaires*. Paris : Gallimard.

Latour, B. (1988). *La vie de laboratoire: la production des faits scientifiques*. Paris : La découverte.

Lerbet, G. (1995). *Les nouvelles sciences de l'éducation*. Paris : Nathan.

Mager, R.F. (1986). *Comment mesurer les résultats de l'enseignement*. Paris : Bordas.

Maingain, A. Dufour, B., Fourez, G. (2002). *Approches didactiques de l'interdisciplinarité*. Bruxelles : De Boeck Université.

Mayen, P. (2005). *Dix leçons sur l'expérience à travers l'expérience de la validation des acquis de l'expérience*. Actes du XVIII<sup>ème</sup> colloque ADMEE. Reims, sept. 2005.

Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : Que Sais-je, PUF.

Monteil, J-M. (1989). *Eduquer et former, perspectives psychosociales*. Grenoble : PUG.

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

Nadeau, M-A. (1988). *L'évaluation de programme, théorie et pratique*. Québec : Les Presses de l'Université Laval.

Parisot, J.-C. (1987). Le paradigme docimologique : un frein aux recherches sur l'évaluation. Cepec. *L'évaluation en questions*. Paris : ESF.

Perrenoud, P. (1995). *La fabrication de l'excellence scolaire*. Genève : Droz.

Rodrigues, P. M. (2006). Logiques d'évaluation en relation avec des conceptions élargies du monde et de la connaissance. In Figari, G., Mottier Lopez, L. (Eds). *Recherche sur l'évaluation en éducation*. Paris : L'Harmattan.

Rodrigues, P. M. (2002). *Avaliação da formação pelos participantes em entrevista de investigação* (= "évaluation de la formation par les participants en entretien de recherche"). Lisboa : Fundação Calouste Gulbenkian.

Scheerens, J. (1989). *Process indicators of school functioning*. Netherlands, University of Twente.

Scheerens, J. (1995). *Internationally comparable indicators of educational programs*. OCDE.

- Shadish, W.R. Jr et al. (1991). *Foundations of Program Evaluation: theories of practice*. (First printing) Newbury Park : Sage.
- Stufflebeam, D.L. et al. (1980). *L'évaluation en éducation et la prise de décision*. Ottawa : NHP.
- Theall, M. (2000). Electronic course evaluation is not necessarily the answer. *Technical Horizons in education*. Source disponible sur : <http://horizon.unc.edu/TS/letters/2000-11.asp>.
- Vinck, D. (2000). *Pratiques de l'interdisciplinarité, mutation des sciences, de l'industrie et de l'enseignement*. Grenoble : PUG.
- Wortman, P.M. (1983). Evaluation research : a methodological perspective. *Annual Review of Psychology*, 34, 223.

**CONFERENCE 1 : LE TESTING ASSISTE PAR ORDINATEUR DANS LA  
MESURE EN EDUCATION: POTENTIALITES ET LIMITES**

**Professeur Romain MARTIN**, Université du Luxembourg

Le testing assisté par ordinateur (TAO) est une pratique de mesure qui existe maintenant depuis plusieurs décennies. Or, malgré des attentes énormes qui étaient formulées dès son introduction vis-à-vis de cette nouvelle méthodologie et de son potentiel innovateur, on doit constater que ces attentes initiales sont loin d'être toutes réalisées jusqu'à ce jour. Suite à une disponibilité de plus en plus importante d'un matériel IT performant dans les institutions de formation et dans les entreprises, ce semble être seulement maintenant qu'une transition à large échelle des instruments classiques de type papier-crayon vers des instruments d'évaluation assistés par ordinateur semble s'amorcer.

Le moment semble donc propice de passer en revue les avantages et les inconvénients de cette technologie d'évaluation, ainsi que les expériences qui ont pu être faites jusqu'à aujourd'hui. Il s'agira également de discuter les possibilités nouvelles offertes par l'ordinateur tant dans le domaine de l'évaluation formative que sommative et qui tiennent avant tout aux possibilités multimédia, aux modalités d'interaction enrichies et aux possibilités d'un traçage comportemental très fin (données chronométriques, enregistrement des interactions comportementales etc.). Ces caractéristiques vont avoir pour conséquence qu'on verra l'émergence de nouveaux dispositifs d'évaluation qui n'étaient pas envisageables avec un format papier-crayon et qui présenteront des tâches dynamiques, éventuellement sous forme de simulations ou de jeux, pouvant impliquer une ou plusieurs personnes. Ces nouveaux dispositifs permettront des évaluations plus centrées sur les processus cognitifs (et autres processus psychologiques) vus sous leur aspect dynamique et émergent et ils ouvriront la porte vers des domaines qui étaient, jusqu'à aujourd'hui, seulement difficilement accessibles à des instruments psychométriques : compétences implicites, compétences sociales, comportement de résolution de problèmes complexes avec haute validité écologique, compétences professionnelles, etc.

Or, au-delà de ces perspectives prometteuses, ces nouveaux dispositifs constituent en même temps un grand défi pour la recherche en évaluation. En effet, ces dispositifs assistés par ordinateur vont engendrer des quantités de données très importantes et nous ne disposons actuellement pas toujours de modèles de mesure adéquats et validés pour procéder à leur traitement. Un autre aspect qui ouvre des perspectives nouvelles pour le TAO est le fait que les ordinateurs dont on dispose aujourd'hui offrent des possibilités de communication par le biais d'une mise en réseau à échelle mondiale à travers Internet. Cette mise en réseau va permettre d'élaborer des outils d'évaluation dans une approche plus collaborative et elle ouvre également des perspectives nouvelles pour la diffusion des tests et pour la centralisation des données. Les possibilités offertes par la mise en réseau pourront notamment être mises à profit dans des programmes de monitoring des systèmes éducatifs qui sont actuellement en

train de se mettre en place dans bon nombre de pays qui veulent évaluer l'efficacité de leurs systèmes de formation à travers les résultats des élèves, souvent dans le contexte de la définition de standards nationaux de formation ou de socles de compétence. Or la mise en place d'une approche collaborative utilisant le TAO dans le contexte de l'implémentation de systèmes de monitoring va également poser la question de la disponibilité d'une plateforme de TAO qui permette de tirer pleinement profit des possibilités de communication que l'outil informatique offre aujourd'hui.

## L'ÉVALUATION DES ÉLÈVES : ENTRE MESURE ET JUGEMENT

Marcel Crahay, Université de Genève

### Mesure et évaluation : de la docimologie à la psychologie de l'évaluation

Comme le rappelle justement Merle (1998), les études docimologiques se sont développées dans l'entre-deux guerre. Elles sont poursuivies jusque dans les années 60 et 70 avec, notamment, la publication de deux ouvrages emblématiques : *Examens et docimologie* de Piéron (1963) et *Evaluation continue et examens. Précis de docimologie* de de Landsheere (1972), réédité à six reprises c'est-à-dire jusqu'en 1980). Cet auteur définit la docimologie comme la « science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés » (1980, p. 13). Cette définition souligne le rôle central occupé dans cette discipline par les examens et, implicitement et plus largement, par les épreuves<sup>1</sup> avec leur corollaire les systèmes de notation. Quant au sous-titre de l'ouvrage de de Landsheere, il reflète par le terme « Précis » le projet normatif de cette science. Celui-ci se nourrit d'un idéal de mesure exacte des performances scolaires. Ceci est manifeste et traité comme une évidence dans les premières éditions. Dans la quatrième édition (1976), les rapports entre mesure et évaluation sont l'objet d'une discussion explicite<sup>2</sup>. Celle-ci repose, d'une part, sur la définition donnée par Guilford (1954) à la mesure et, d'autre part, sur un texte de Taba (1962, cité par de Landsheere, 1976)) distinguant mesure et évaluation. Pour ce dernier, « l'évaluation dépend de la mesure, mais elle porte sur un profil plus large de caractéristiques et de performances » (p. 17). Dénonçant le postulat implicite consistant à supposer que toute évaluation repose une quantification, de Landsheere (1976) évoque l'évaluation d'une œuvre d'art pour conclure que « bien des comportements et combien d'œuvres humaines s'évaluent de pareille façon » (p. 17). Selon notre lecture du Précis de docimologie, cette argumentation substitue un dualisme opposant l'évaluation qualitative à la mesure quantitative en lieu et en place d'une dépendance de la première par rapport à la seconde telle que défendue par Taba (1962). Bref, d'un côté, on trouve l'évaluation qui relève de l'appréciation quasi esthétique des choses et, de l'autre, la mesure dont de Landsheere (1976, p. 16) écrit, en suivant Guilford (1973) qu'elle est une opération consistant à « assigner un nombre à un objet ou à un événement selon une règle logiquement acceptable ». Ceci implique que la mesure porte sur des objets clairement circonscrits et qu'une règle permette de lier systématiquement les valeurs métriques aux variations des objets. Le reste de l'ouvrage, quasi inchangé par rapport aux versions précédentes, perpétue le projet docimologique : fonder une science de la notation qui soit aussi objective que possible.

Le projet qui anime de Landsheere depuis 1970 (première version) est noble. Il s'agit de juguler les biais multiples qui introduisent l'inégalité dans les notations

---

<sup>1</sup> Rappelons que la racine grec – dokimè – du terme docimologie signifie « épreuve ».

<sup>2</sup> On ne trouve pas trace de cette discussion dans la deuxième édition qui date de 1972. Nous ne possédons pas la troisième.

scolaires des enseignants. Ce projet n'est ni neuf (cf. parmi d'autres, Piéron, 1963), ni isolé (Bonboir, 1972). Il n'est pas exempt de contradictions. Ainsi, introduisant un important chapitre consacré à la notation, de Landsheere (1976) traite de la courbe de Gauss comme un préambule indispensable (p. 116) avant d'en dénoncer les méfaits dans un paragraphe intitulé « le dangereux mythe de la courbe de Gauss » (cf. p. 210). Pour lever cette contradiction, il convenait d'abandonner l'ancrage psychométrique et rechercher un autre arrimage axiomatique pour fonder l'éducativité. Ce fût l'œuvre de Cardinet (1988a, b) et de ses études sur la généralisabilité. Nous y reviendrons dans la section suivante.

Parallèlement se développe une autre approche, résolument enracinée dans la psychologie. Elle repose sur l'idée de base selon laquelle « l'activité d'évaluation fait partie intégrante de nos modalités d'adaptation à l'environnement physique et social » (Aubret, 1995, p. 541). Bref, il s'agit de reconnaître que l'évaluation scolaire mobilise des opérations mentales de traitement de l'information et, donc, des activités perceptives et cognitives. Selon Noizet et Caverni (1976), le noyau de ces opérations procède d'un schéma de comparaison, consistant à rapporter un objet à évaluer sur une échelle de valeur. Plus précisément, tout évaluateur sélectionne des observables qu'il va interpréter en fonction de systèmes de référence. Ainsi, évaluer le savoir lire, c'est sélectionner certains indices ou indicateurs parmi une très grande diversité d'observables possibles (vitesse de lecture, fluidité de la lecture à haute voix, erreurs et hésitations, façons de reformuler les idées du texte, etc.) et les interpréter en termes de niveau. Le système de référence de l'individu intervient d'emblée dès l'étape de sélection des observables et influe également sur l'interprétation de ceux-ci.

Par ailleurs, les performances scolaires ne sont pas dissociables des élèves qui les produisent. Il est donc légitime de s'interroger, comme le fait Gilly (1980) sur les mécanismes sous-jacents aux jugements scolaires portés par les enseignants sur les élèves. Selon cet auteur, l'enseignant perçoit l'élève à travers un cadre schématique qui est à la fois sélectif, organisateur et interprétant des informations disponibles. Or, ce cadre est façonné par les objectifs professionnels et les normes socio-institutionnelles qui habitent chaque enseignant. Plus précisément, les jugements des enseignants sont le produit des rapports entre trois déterminants : on trouve, à un premier niveau, l'expérience quotidienne (activités, indices comportementaux, etc.), au niveau intermédiaire, les représentations sociales et, à l'autre extrémité, les normes sociales générales (valeur morale, modèle de l'homme dans la société, etc.) et les normes scolaires institutionnelles (règles de fonctionnement de l'école, objectifs, etc.). Par conséquent, la perception de l'élève par l'enseignant est socialement et institutionnellement déterminée.

### Évaluation et éducatrice : les apports de Cardinet

Dès la fin des années 70, se dessinent les axes selon lesquels vont s'organiser la réflexion et les travaux de recherche. Le titre de l'ouvrage édité par De Ketele en 1986 révèle un de ces axes : *L'évaluation : approche descriptive ou prescriptive ?* Le second axe se révèle à la lecture des titres donnés à deux ouvrages de Cardinet : *Evaluation scolaire et mesure* (1988a), pour le premier ; *Evaluation scolaire et pratique* (1988b), pour le second. D'une part, il s'agit à la fois de comprendre le fonctionnement des enseignants et, parallèlement, de les conseiller pour aboutir à des évaluations d'élèves plus justes. D'autre part, il s'agit d'assumer le débat qualitatif *versus* quantitatif dans le champ de l'évaluation scolaire.

Élève de Cronbach, Cardinet est d'abord un homme de mesure, ce qui ne l'empêche pas de se montrer soucieux des questions d'équité éducative. Bien plus que de Landsheere, il assume les perversités de la loi de normalité statistique énoncée par Gauss en matière d'évaluation pédagogique. Prolongeant le travail de Cronbach, Gleser, Nanda et Rajaratnam (1972), il développe, en collaboration avec Tourneur (Cardinet & Tourneur, 1985), la théorie de la généralisabilité et, ce faisant, contribue à l'émergence de l'éducatrice.

Née dans le contexte des théories eugénistes (Lemaine & Matalon, 1985), la psychométrie classique se donne pour objet de discriminer de façon fiable les individus, principalement en fonction de leurs compétences cognitives (leurs aptitudes). Dans cette perspective, la mesure est jugée fidèle si les résultats pour les différentes questions du test sont assez semblables ; techniquement, il s'agit de s'assurer que la variance inter-questions soit inférieure à la variance entre individus. La mesure obtenue peut être attribuée à l'individu puisqu'elle est indépendante des fluctuations du paramètre « question ». Cependant, objecte Cardinet (1988a), « cette conception de la fidélité, (...) ne permet pas d'estimer la fidélité de la mesure d'élèves pris isolément ». Il convient donc de chercher le moyen d'estimer la compétence de l'élève « sans du tout faire intervenir les résultats d'autres personnes » (p. 121). Plus largement, il s'agit d'assumer, sur le plan de la mesure, les différentes questions qui se posent en éducation. Certes, il faut continuer à pouvoir différencier les individus, notamment dans la perspective de composer des groupes de besoin ou des groupes coopératifs, mais il faut aussi pouvoir distinguer les domaines de compétences au sein desquels l'élève fait preuve de maîtrise de ceux au sein desquels il éprouve des difficultés et ceci dans la perspective de diagnostiquer ses forces et ses faiblesses. Il faut encore pouvoir non seulement examiner les scores d'un élève en fonction des moments de l'évaluation et ceci afin de faire apparaître ses évolutions, mais aussi envisager la possibilité de relations hiérarchiques entre les diverses compétences évaluées.

En contact régulier avec le monde des enseignants, Cardinet prend conscience des limites des approches scientifiques classiques. Dans un article intitulé *L'élargissement de l'évaluation* (1979), il remet en cause l'ancrage positiviste des procédures classiques des mesures psychologiques et pédagogiques et ouvre le champ de l'évaluation aux approches compréhensives. Dans la foulée, il s'interroge « faut-il encore mettre des notes ? » (1988a). Répondant de façon négative, il pose alors la question « pour évaluer les élèves, que faut-il observer ? » (1988b). Plus largement, son ouvrage *Evaluation scolaire et pratique* s'efforce d'outiller conceptuellement la

pratique de l'évaluation en prenant en charge ses diverses fonctions : pronostique, diagnostique-formatrice et sommative/certificative. Or, il n'est pas légitime d'utiliser la même mesure pour ces diverses fonctions. Il convient de recueillir des informations spécifiques en fonction de la question posée. Ainsi, pour choisir une démarche correctrice, il semble pertinent non seulement de cerner ce que l'élève maîtrise et de cerner ses lacunes et difficultés, mais aussi d'appréhender comment cet élève a structuré les fragments de connaissances maîtrisées et même de retracer l'historique ou le cheminement de l'apprentissage afin d'identifier le (ou les) moment(s) du processus où le (ou les) dérapage(s) s'est (ou se sont) produit(s).

### **Jugement scolaire et décision de redoublement**

Désormais le terrain de l'évaluation formative semble être réservé aux approches qualitatives. Celui de l'évaluation sommative et/ou certificative est plus problématique. En effet, les épreuves communes ou externes suscitent bien des réactions de la part des enseignants et certaines de leurs objections sont légitimes. Nombre de questions composant ces épreuves ne sont pas - disent-ils - en adéquation avec l'enseignement qui a été offert aux élèves ; il n'est donc pas étonnant que ceux-ci soient désorientés et, perdant pied pour certains, se montrent peu performants. Bref, se pose une question de validité : que mesure-t-on par ces épreuves ? Les effets de l'enseignement reçus ou la capacité des élèves à s'adapter à l'imprévu ? Pourquoi, dès lors, ne pas accepter que les enseignants soient seuls habilités à concevoir les épreuves sommatives adaptées à leur classe, voire même à décider, sur la base de leurs perceptions quotidiennes des progrès de chaque élève ceux qui peuvent passer de classe et ceux qui doivent redoubler ? De facto, ces arguments tendent à légitimer l'état de la situation. Une situation, dont certains chercheurs ont montré les conséquences néfastes en matière d'échec scolaire : l'idiosyncrasie des enseignants produit des inégalités flagrantes.

Ainsi, en Communauté Française de Belgique, Grisay (1988) a pu montrer qu'à performances égales à une épreuve externe, certains élèves réussissent dans l'école X tandis que d'autres échouent dans l'école Y. L'analyse des épreuves conçues par les enseignants fait apparaître que le niveau de difficulté des questions varie de façon considérable et, partant, que les critères sur lesquels les enseignants se basent pour décider de la promotion ou du redoublement fluctuent. En fait, les enseignants ajustent leurs épreuves au niveau moyen du groupe d'élèves : classe jugée forte, exigences accrues ; classe jugée faible, mansuétude plus grande. Ce mécanisme d'adaptation a pour effet que, dans quasiment toutes les classes, deux élèves sont chaque année en échec. Interrogés à ce sujet, les enseignants se déclarent conscients de cet ajustement qu'ils qualifient de différenciation pédagogique.

Par ailleurs, ré-analysant des protocoles d'entretien recueillis par Burdevet (1996) à Genève, Crahay (2003) fait apparaître qu'au moment de décider le passage de classe ou le redoublement, les enseignants troquent leur rôle d'activateur des apprentissages pour celui de psychologue. En effet, la décision de promotion ou d'échec ne se fonde guère sur un bilan des acquisitions réalisées durant l'année écoulée. Les enseignants s'efforcent de pronostiquer dans quelle mesure chaque élève est « apte à suivre les enseignements de l'année à venir ». Pour justifier leurs verdicts d'incapacité à l'adresse de certains enfants, les enseignants invoquent principalement des causes non

scolaires : manque de maturité et, plus rarement, d'aptitude, habileté langagière déficiente, problèmes affectifs, mais aussi famille immigrée ou perturbée. Tous les enseignants sans exception affirment que les notes ne constituent pas l'élément déterminant d'une décision de redoublement. Ce qui importe, déclarent-ils, c'est leur jugement auquel la plupart accordent une grande confiance à la limite de l'inaffabilité. Plus précisément, bon nombre des enseignants interrogés semblent considérer que leurs jugements sont objectifs du fait même qu'ils sont en contact quotidien avec les enfants. Inconscients de possibles effets d'attentes, certains avouent toutefois être influencés par la représentation qu'ils se font des exigences du collège de l'année ultérieure. Quant aux notes, elles constituent « une sorte d'alibi » (propos d'enseignant) en ce sens qu'elle a pour fonction de donner à la décision prise une légitimité aux yeux de l'élève, des parents et des collègues.

### **Réconcilier le jugement et la mesure**

L'époque de la toute-puissance de la mesure est révolue. Ses limites ont été cernées avec justesse par Cardinet (1988a) qui, par ailleurs, a développé, au travers de la théorie de la généralisabilité, les fondements d'une éducatrice libérée de la courbe de Gauss. Il est urgent que la corporation enseignante fasse la critique de l'évaluation par jugement. Les travaux de Bressoux et Pansu (2003) sur les jugements scolaires offrent une base empirique et théorique à cette entreprise.

Selon nous, il est indispensable de reconnaître les limites tant des mesures éducatrices que des jugements scolaires des enseignants. C'est à ce prix qu'il sera possible de concevoir des dispositifs d'évaluation des élèves combinant des mesures, certaines provenant d'épreuves spécifiques à la classe et d'autres issues d'épreuves externes, et les jugements des enseignants. La base de l'appréciation des élèves restera sans doute le jugement des enseignants, mais ceux-ci doivent faire le deuil de leur foi en l'inaffabilité de leurs perceptions afin de fonder leurs décisions certificatives sur une pluralité d'informations et accorder un poids véritable aux scores des élèves à diverses épreuves. Le but est, en définitive, d'assumer l'incertitude inexorablement liée à l'appréhension des qualités humaines.

**Références bibliographiques**

- Aubert, J. (1995) Psychologie de l'évaluation, (pp. 540-555) In D. Gaonach, D. & C. Golder, (Eds.). *Profession enseignant : manuel de psychologie pour l'enseignement*. Paris : Hachette.
- Bonboir, A. (1972). *La docimologie* Paris : Presses Universitaires de France.
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves* Paris : Presses Universitaires de France.
- Cardinet, J. (1979). L'élargissement de l'évaluation. *Education et recherche*, 1 (1), 15-34.
- Cardinet, J. (1982). *Pour évaluer les élèves, que faut-il observer ?* Neuchâtel, Rapport SIPRI, Cahier 4, IRDP. (PAS CITE)
- Cardinet, J. (1985). Des instruments d'évaluation pour chaque fonction. *Mesure et évaluation en éducation*, 8(1-2), 45-116. (PAS CITE)
- Cardinet, J. (1988a). *Évaluation scolaire et mesure* Bruxelles : De Boeck.
- Cardinet, J. (1988b). *Évaluation scolaire et pratique* Bruxelles : De Boeck.
- Cardinet, J. & Tourneur, Y. (1985). *Assurer la mesure : guide pour les études de généralisabilité*. Berne, Peter Lang.
- Chevallard, Y. (1986). Vers une analyse didactique des faits d'évaluation. In J. M. D. Ketele (Ed.), *L'évaluation : approche descriptive ou prescriptive*. Bruxelles, De Boeck, Collection "Pédagogies en développement". (PAS CITE)
- Crahay, M. (2003). *Peut-on lutter contre l'échec scolaire ? pratique* Bruxelles : De Boeck (2<sup>e</sup> édit).
- Cronbach, L. Gleser, G. Nanda H. & Rajaratnam, N. (1972), *The dependability of behavioral measurement : theory of generalizability for scores and profiles*. New York : John Wiley.
- De Ketele, J.-M. (Ed.). (1986), *L'évaluation : approche descriptive ou prescriptive ?*. Bruxelles : De Boeck.
- de Landsheere, G. (1972). *Evaluation continue et examens. Précis de docimologie* Bruxelles : Labor, 1972.

- Gilly, M. (1980) *Maître-élève. Rôles institutionnels et représentations*. Paris : Presses Universitaires de France.
- Grisay, A. (1988). *Du mythe de la "bonne école" à la réalité (fuyante) de l'"école efficace"*. Liège, Service de Pédagogie expérimentale de l'Université.
- Guilford, J.P. (1954). *Psychometrics Methods*. New York : Mc Graw (2<sup>e</sup> édit).
- Lemaine, G. & Matalon, B. (1985), *Hommes supérieurs, hommes inférieurs ? La controverse sur l'hérédité de l'intelligence*. Paris : Armand Colin.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire* Paris : Presses Universitaires de France, Que sais-je ? 3278.
- Noizet, G. & Caverni, J.P. (1976). *Psychologie de l'évaluation scolaire* Paris : Presses Universitaires de France.
- Piéron, H. (1963). *Examens et docimologie*, Paris : Presses Universitaires de France.

**ETAT DES LIEUX ET PERSPECTIVES POUR L'ÉVALUATION DE LA  
QUALITÉ DE L'ENSEIGNEMENT SUPÉRIEUR EN EUROPE**

**Fabrice Hénard**, chargé de mission au Comité national d'évaluation (CNÉ)

**Mots-clefs** : qualité, évaluation, enseignement supérieur, convergence.

**Résumé :**

Cet article reflète l'importance prise par l'évaluation de l'enseignement supérieur en Europe et l'émergence de la notion de qualité. Il souligne la diversité des objectifs assignés aux agences d'évaluation et le rapprochement des méthodes d'évaluation. Il conclut sur les perspectives de convergence européenne.

Cet article a été rédigé à partir de contributions et présentations existantes, qui figurent en annexe. Il ne reflète pas un point de vue particulier du CNÉ et ne prétend pas non plus à montrer l'exhaustivité des réflexions nombreuses en Europe sur ce sujet évolutif. Son objectif est de synthétiser la situation telle qu'elle est en juillet 2006.

---

**I - La dimension européenne de l'évaluation de l'enseignement supérieur**

**De l'évaluation de l'enseignement supérieur en Europe**

L'évaluation n'est pas une idée neuve, les systèmes d'enseignement supérieur se sont toujours souciés de la qualité de leurs activités de formation et de recherche. La nouveauté provient du fait que la demande d'évaluation se précise et se diversifie. Les États, les multiples financeurs de l'enseignement supérieur, les étudiants, les employeurs et les universités elles-mêmes formulent le besoin de mieux connaître les activités des universités et d'en mesurer la performance. L'expression de la demande d'évaluation reste plus ou moins explicite selon les acteurs, et opère volontiers un glissement sémantique entre évaluation et qualité. Les multiples textes qui traitent d'évaluation de la qualité, d'assurance qualité ou de management de la qualité de l'enseignement supérieur précisent rarement le sens de ces notions (Dejean, actes du colloque du CNÉ à Dijon, 2004).

Aujourd'hui, l'évaluation de la qualité est un concept communément partagé par tous les systèmes d'enseignement supérieur du monde entier. Elle apparaît comme un moyen permettant de comprendre la complexité de l'enseignement supérieur et de garantir la qualité de l'enseignement prodigué. L'évaluation de la qualité fait l'objet de nombreuses réflexions et de colloques nationaux ou internationaux. Force est de reconnaître, cependant, que la matière demeure encore relativement inexplorée dans toutes ses dimensions, et particulièrement à l'échelle européenne. Il reste délicat de définir l'évaluation de la qualité de l'enseignement supérieur et de saisir les acceptions qu'elles recouvrent. Les acteurs de l'enseignement supérieur manquent

encore de recul sur l'évolution historique et le cadre philosophique et politique de l'évaluation de la qualité.

L'approche européenne de l'évaluation de l'enseignement supérieur ne peut être exhaustive tant est grande la diversité des systèmes nationaux d'enseignement supérieur. Chaque système est effectivement mû par un nombre important de facteurs nationaux qui ont leurs propres dynamiques : la démographie étudiante, les enjeux des territoires, les besoins de recrutement, les calendriers politiques... (*Quality convergence study*, ENQA, Crozier, Curvale, Hénard, 2005).

Il semble clair qu'un mouvement de convergence se dessine au niveau européen, tandis que parallèlement les politiques nationales d'enseignement supérieur s'affirment davantage. De cette tension naît une confusion sur le sujet de l'évaluation de la qualité en Europe. Mais cette confusion aujourd'hui apparente n'est peut-être, en réalité, que le reflet d'une diversité encore mal décryptée. Les chercheurs proposeront sans doute dans l'avenir les clefs de compréhension qui permettront d'appréhender plus clairement la situation. Il convient donc pour le moment de rester prudent sur le sujet et d'accepter que les idées et les réflexions, loin d'être figées dans le temps, sont destinées à évoluer rapidement.

Depuis maintenant près de 10 ans, la construction de l'espace européen de l'enseignement supérieur offre un cadre nouveau à l'évaluation de la qualité. Dès lors que l'espace européen de l'enseignement supérieur se construisait, il lui devenait nécessaire de se doter des outils pour l'évaluation de la qualité et de montrer comment la qualité était assurée par les établissements. Certains auteurs considèrent même que l'assurance qualité est devenue une véritable politique publique en Europe (Perellon, 2004). Parallèlement, les objectifs de l'évaluation se sont progressivement élargis au fur et à mesure du développement des échanges interuniversitaires en Europe, et de l'avancée rapide de processus politiques intégrateurs, au premier rang desquels figure le processus intergouvernemental de Bologne (Weber, Bologna Handbook, 2006).

Un consensus est apparu entre les États signataires du processus de Bologne sur l'indispensable dimension européenne de l'évaluation de l'enseignement supérieur. Il considère que l'isolement national des systèmes d'évaluation risquerait de les enfermer dans un cadre restreint, coupé des réalités de la "société de la connaissance" qui s'ancre sur des objets transnationaux (Haug, 2003). La recherche, et maintenant la formation et l'insertion professionnelle s'internationalisent. Par effet de miroir, l'évaluation doit s'y adapter.

Élargi à d'autres États non membres de l'Union européenne, le processus de Bologne devient un puissant levier d'harmonisation de l'enseignement supérieur, en dehors de tout cadre contraignant, de tout financement et de tout instrument de contrôle qui s'imposerait à tous. Ce processus a progressivement introduit l'évaluation de la qualité de l'enseignement supérieur dans ses objectifs, au point qu'elle en devienne aujourd'hui un des piliers (Kahn, 2003 ; Froment, Bologna Handbook, 2006).

À Bologne, en 1999, la déclaration évoque la mise en oeuvre de la coopération européenne dans le domaine de l'assurance qualité avec l'objectif d'instaurer des critères et des méthodes comparables. En 2001, lors de la conférence de Prague, les ministres prônent la coopération européenne en matière d'assurance qualité, insistent sur l'importance des systèmes nationaux d'assurance qualité et soutiennent l'émergence d'un cadre commun de réflexions sur le sujet pour repérer les bonnes pratiques. À partir de la conférence de Berlin en 2003, les ministres s'engagent à soutenir le développement de l'assurance qualité à la fois dans les établissements et sur une échelle nationale et européenne. Ils soulignent la nécessité de définir des critères et des méthodologies mutuellement reconnus. Ils tiennent à rappeler également que, conformément au principe de l'autonomie des établissements, la responsabilité première de la qualité incombe à chaque établissement, ce qui fonde les bases d'une réelle responsabilisation du système universitaire dans le cadre national. Ils reprennent en cela l'idée de l'Association européenne des universités (EUA) émise à Graz en 2003.

À Bergen, en 2005, des références et lignes directrices européennes d'assurance qualité sont adoptées, le modèle d'évaluation des agences par les pairs sur une base nationale est introduit, le principe d'un registre européen des agences d'assurance-qualité est posé et l'importance de la coopération entre agences nationalement reconnues est affirmée.

L'Union européenne porte une autre dimension européenne de l'évaluation. Engagée depuis plusieurs décennies dans des programmes visant à favoriser la mobilité, dont ERASMUS, l'Union européenne s'est progressivement intéressée à l'éducation et à l'enseignement supérieur, secteurs qui relèvent cependant de compétences nationales (Office des publications des communautés européennes [OPCE], 2006). Mais, initiatrice de la stratégie dite de Lisbonne visant à faire de l'économie européenne la plus compétitive au monde en 2010, l'Union tend à considérer que les systèmes éducatifs sont des facteurs décisifs de la croissance et de la compétitivité économique dans le monde globalisé. La stratégie de Lisbonne, ainsi que d'autres processus de l'Union, comme celui de Copenhague qui promeut un cadre commun pour la formation professionnelle, croise quelque part les intérêts du processus de Bologne (Froment, Bologna Handbook, 2006) sans toutefois les confondre. En outre, l'espace européen de l'enseignement supérieur s'articule naturellement avec l'espace européen de la recherche (la recherche est une compétence de l'Union partagée avec les États membres), ces espaces réunissant des préoccupations proches en évaluation de la qualité.

Mais ces liens entre les différents processus ne sont pas exempts de tensions. Par exemple, la dimension sociale du processus de Bologne impliquerait la recherche d'une cohésion sociale envers l'étudiants dans l'espace européen, tandis que la stratégie de Lisbonne de l'Union viserait plutôt l'excellence universitaire pour servir des objectifs de compétitivité économique.

Malgré cela, un des points d'accord aujourd'hui est que la dimension européenne de l'évaluation de l'enseignement supérieur ne conduit pas à un système unique, européen, d'évaluation. Cet idée est dorénavant abandonnée, selon toutes hypothèses. Les États ont sans doute considéré que les établissements d'enseignement supérieur étaient trop divers et les objectifs de l'évaluation trop nombreux pour envisager un système d'évaluation uniforme valable pour l'ensemble de l'Europe. Par ailleurs, les gouvernements accordent un intérêt croissant à l'enseignement supérieur et à son impact sur les processus d'innovation et sur les économies en général ; ceci les conduit à s'interroger sur l'appréciation de la qualité des établissements et à engager des réformes législatives et réglementaires pour améliorer le volet évaluation de leur système national d'enseignement supérieur.

## **II - La diversité des objectifs de l'évaluation de l'enseignement supérieur en Europe et la similarité des méthodes**

### **Une multitude de missions au sein même de chaque agence d'évaluation**

À l'examen des missions des agences chargées d'évaluer l'enseignement supérieur que propose l'étude conduite par l'ENQA (*ENQA occasional papers 5*, the Danish Evaluation Institute, 2003), on est frappé par leur diversité. Les agences remplissent trois missions principales qui peuvent se conjuguer : l'amélioration de la qualité dans les établissements d'enseignement supérieur, la diffusion de connaissances et d'informations auprès des étudiants, des employeurs ou de la société en général et l'accréditation de formations ou d'établissements. Mais deux autres types de missions sont apparus sous les pressions de la construction de l'Europe de l'enseignement supérieur : favoriser la transparence des activités des universités et rendre compte de ce que produit le système d'enseignement supérieur (en référence à la notion d'*accountability*). Les classements et les notations demeurent, en revanche, peu prisés des agences d'évaluation.

L'accréditation et l'évaluation de programme sont les deux activités les plus fréquemment exercées.

**Tableau 1** – Les activités des agences des pays signataires du processus de Bologne<sup>1</sup>

	Évaluation	Accréditation	Audit	Benchmarking
Sujet/discipline	6	1	1	6
Programme	21	20	15	7
Établissement	12	10	14	4
Thème	10	0	1	4

Source : *Quality procedures on higher education, an ENQA survey, 2003*

*L'évaluation de discipline* cible la qualité de celle-ci dans le cadre de tous les programmes qui la proposent (par exemple, la chimie organique dans un Master de sciences de la matière). *L'évaluation de programme* cible les activités d'un programme qui conduisent à l'obtention d'un diplôme. *L'évaluation de programme* est prisée des agences des pays scandinaves, du Royaume-Uni et des Pays-Bas, ainsi que des agences opérant dans le secteur non-universitaire, caractérisé par sa forte professionnalisation. En revanche, peu d'agences pratiquent l'évaluation de disciplines. *L'évaluation institutionnelle* examine l'ensemble des activités de l'établissement (organisation, finances...). Près de la moitié des agences la pratique en Europe. *L'évaluation de thème* examine la qualité ou la pratique d'un thème spécifique dans un cadre éducatif, par exemple les TIC ou le tutorat d'étudiants.

*L'accréditation* est une autorisation accordée à un établissement de pratiquer, qui est fondée sur la satisfaction d'un niveau minimum prédéterminé. *L'accréditation* est conçue comme un moyen de garantir la qualité des formations dans le souci de protéger les étudiants, mais aussi de favoriser l'amélioration de la qualité par l'instauration de normes qui permettent d'établir une qualité normale ou de base. Elle doit permettre de protéger le terme même d'*université*, particulièrement dans les pays où le secteur universitaire privé explose (par exemple dans les pays de l'Est et de

<sup>1</sup> Sur la base de 34 agences pour 23 pays en 2003

l'Europe orientale). Elle est caractérisée par le fait qu'elle conduit à une approbation officielle d'un établissement ou d'un programme. Elle se manifeste concrètement sous la forme d'une décision basée sur le respect ou la conformité avec des exigences ou des standards prédéfinis (qui sont en général les normes minimales), officielle, de nature binaire (oui/non) ou tertiaire (sous conditions), valide pour un temps limité.

Il est fréquemment considéré que la mise en œuvre de procédures d'accréditation exige un découplage entre l'avis de l'évaluateur académique ou scientifique et la décision politique prise généralement par le ministre de l'enseignement supérieur.

Les perspectives ultimes de l'accréditation et de l'évaluation ne sont pas les mêmes. Dans le cas de l'évaluation, la responsabilité de la qualité est, de manière pleine et entière, celle du maître d'œuvre de la formation. Dans le cas de l'accréditation, la responsabilité de la qualité passe par le respect d'un cahier des charges relativement prescriptif. Mais, en réalité, l'exigence de conformité que sous-tend l'accréditation est souvent confondue avec la recherche de l'amélioration de la qualité. Le débat reste ouvert pour savoir si l'accréditation, dont la vocation est d'aboutir à une décision de type "oui ou non", remplit bien les missions qu'on assigne généralement aux dispositifs d'évaluation de la qualité. Ce débat, en Europe, n'est guère facilité par la terminologie, mouvante d'un pays à l'autre et dans le temps. Aujourd'hui, des interrogations importantes demeurent, concernant par exemple la légitimité de l'accréditeur, la nature de l'engagement de sa responsabilité et de celle de l'autorité qui décide, la pérennité de l'autorisation accordée, le renouvellement de l'accréditation...

Il existe de nombreux terrains où évaluation et accréditation se rejoignent pleinement et il serait artificiel d'opposer les deux activités. Les accréditeurs et les autres évaluateurs reconnaissent que le contrôle de la conformité à des normes de qualité minimale et que la nécessité de favoriser l'amélioration continue de la qualité sont légitimes. Les accréditeurs tendent à employer des méthodes d'évaluation et substituent à un avis binaire (accréditation ou refus d'accréditation) des niveaux de recommandation gradués dans le but d'inciter l'université à s'améliorer. Pour leur part, les évaluateurs introduisent des notions de comparaison inspirées des pratiques d'accréditation.

*L'audit* est une méthode d'évaluation qui évalue les forces et faiblesses des mécanismes d'assurance qualité que les établissements ont eux-mêmes mis en place pour leurs propres besoins, afin de piloter et d'améliorer les activités et les services associés à une discipline, à un programme, à un établissement ou à un thème. L'audit n'est pas très répandu en Europe continentale, hormis au sein des agences britannique et irlandaise et de quelques agences nordiques.

*Le benchmarking* est une méthode comparative de résultats entre les disciplines, les programmes, les établissements ou les thèmes, dans le but de repérer les meilleures pratiques. À la différence de l'accréditation, qui s'appuie généralement sur des critères minimum de qualité, le benchmarking vise l'excellence. Le

benchmarking d'établissement est très rare et reste peu développé pour les disciplines et les programmes.

### **Les avancées inégales dans la mise en place de systèmes d'assurance qualité en Europe**

Une cartographie des pratiques et des objectifs d'évaluation aussi variés est délicate à mener en Europe, d'autant plus que leurs évolutions sont permanentes. Dans presque tous les pays signataires du processus de Bologne, l'évaluation est un sujet en discussion et dont les objectifs, les pratiques et les méthodes changent régulièrement.

L'avancée du processus de Bologne permet toutefois de confirmer que la grande majorité des pays signataires ont progressé dans le développement de leur système d'évaluation de l'enseignement supérieur et intègrent dans leurs stratégies l'obligation de se conformer aux engagements pris à Berlin en 2003 et à Bergen en 2005 (Bologna stocktaking, 2005).

Presque tous les signataires du processus de Bologne disposent d'un système opérationnel d'assurance qualité. La définition des responsabilités entre les agences et les établissements d'enseignement supérieur est jugée satisfaisante. La plupart des systèmes d'assurance qualité se limitent toutefois à la mise en place d'une agence nationale ou régionale indépendante. D'autres pays se sont appuyés sur des agences existantes et ont étendu leurs missions pour évaluer la qualité. Dans quelques pays encore, la mission reste exercée par un service du ministère ou une agence gouvernementale.

Le processus de Bologne a défini cinq éléments clés d'un système d'évaluation : l'évaluation interne, l'évaluation externe, la participation des étudiants, la publication des résultats et la coopération internationale sur la qualité. En 2005, dans 18 pays, ces cinq éléments sont totalement acceptés et mis en place. En revanche, la participation des étudiants est l'élément clé le moins avancé et, bien souvent, les structures de participation ou de représentation étudiante ne sont pas suffisamment développées. La publication des rapports d'évaluation n'est pas non plus toujours acceptée, ce qui pose le problème de la transparence des résultats d'évaluation. Quant au niveau de coopération internationale, il est estimé correct puisque toutes les agences s'investissent dans les organisations européennes et internationales traitant d'évaluation et d'assurance qualité.

### **Une certaine similarité des méthodes d'évaluation**

Malgré leur diversité, les méthodes d'évaluation de l'enseignement supérieur se rapprochent. Les chercheurs Lee Harvey et Jethro Newton (2004) indiquent que les méthodes dépendent des objectifs et de l'intérêt de l'évaluation selon les prestataires de formation (les établissements), les apprenants (les étudiants), les gouvernants (les

États et autres financeurs), les utilisateurs de formation (les employeurs). L'évaluation conduit également à choisir une méthode plutôt qu'une autre selon qu'elle cherche à évaluer la gouvernance et la gestion, l'élaboration des programmes, leur contenu, la capacité d'apprentissage, le mode de délivrance des diplômes, l'aide à l'étudiant, la viabilité financière, la qualification, le support administratif, l'organisation et les procédures. Le contexte de l'évaluation influe également sur la méthode selon que l'évaluateur souhaite rendre des comptes, contrôler, améliorer ou faire respecter des standards de qualité.

Plusieurs méthodes existent (Williams, 2006) : l'évaluation par les pairs, l'expertise, l'utilisation de critères ou de standards, l'analyse par indicateurs de performance, les panels d'experts externes ou internes à l'établissement, l'analyse documentaire, les enquêtes par ou sur les parties prenantes, les études de cas, les observations directes de pratiques. Lee Harvey évoque "l'actuelle méthodologie dominante que sont l'évaluation interne, l'évaluation par les pairs et l'analyse documentaire et statistique, accompagnée par la publication du rapport d'évaluation" (Harvey, 2004, p 70). Sans passer en revue toutes les méthodes, il est intéressant de se pencher sur celles qui prévalent actuellement dans l'évaluation de l'enseignement supérieur, comme le propose l'*ENQA occasional papers 5* (2003).

#### *L'évaluation interne ou auto-évaluation*

Dès lors que le but premier des agences est l'amélioration de la qualité et que celle-ci relève d'abord de la responsabilité des établissements, il n'est pas surprenant que l'auto-évaluation soit devenu le pivot des procédures d'évaluation. L'utilité et l'usage de l'auto-évaluation dépendent essentiellement des contextes nationaux et des habitudes des universités. Dans les pays où la culture de l'évaluation est répandue, l'auto-évaluation est un exercice normal de l'activité. De même, les établissements fortement concurrentiels y souscrivent souvent sans difficultés.

La pratique courante au niveau européen, reflétée dans les références et lignes directrices européennes, est de considérer l'auto-évaluation comme un moyen laissé aux établissements pour analyser leur situation sur la base d'arguments. La logique de démonstration supplante la logique de description. On attend de l'établissement qu'il soit capable de démontrer qu'il est responsable de ses activités.

#### *L'évaluation par les pairs*

L'évaluation par les pairs est une dimension de l'indépendance de l'évaluation et de sa crédibilité. L'universitaire est naturellement qualifié pour évaluer une activité universitaire. Les références européennes précisent que les experts choisis doivent réunir les compétences adaptées à la mission qui leur est confiée, que l'agence doit faire preuve de transparence dans la sélection de ses experts et qu'il doit être fait appel à des experts internationaux ainsi qu'à des étudiants.

Cette méthode devient objet de débat. En effet, partout en Europe, l'évaluation se complexifie en intégrant une dimension internationale et en ciblant davantage le management de la qualité de l'établissement. Elle inclut des étudiants de manière plus systématique, elle aborde de nouveaux sujets d'évaluation qui dépassent les frontières des disciplines académiques (par exemple, la relation université-entreprise). Elle nécessite donc des experts ayant une connaissance profonde de l'université et des disciplines, mais étant également pourvus d'une capacité d'analyse systémique et stratégique.

Par conséquent, le recrutement des experts et le pilotage de l'équipe d'experts devient une activité cruciale pour les agences.

Sur le choix des experts internationaux, la situation diffère selon les types d'évaluation. L'agence d'évaluation dont l'objectif est d'aider l'université à s'améliorer emploie sans difficultés des experts internationaux, qui apportent un regard neuf sur l'appréhension de situations nationales. Les agences d'accréditation, dont l'avis conduit à une décision d'autorisation ou de refus ou conditionne un financement, prennent certains risques à recruter des experts internationaux parfois moins au fait des contextes nationaux et des subtilités du fonctionnement du système d'enseignement supérieur. Le recrutement d'experts internationaux pose également la question de la langue, notamment pour les pays non anglophones, ce qui peut conduire à des échanges d'experts limités à une même aire culturelle. Le recrutement d'experts étudiants semble plus simple, bien que leur représentativité fasse débat. Les étudiants choisis doivent-ils être représentatifs de la population étudiante de l'établissement évalué ? Concernant les experts issus du monde socio-économique, leur recrutement est une pratique encore assez peu répandue alors qu'ils sont d'une aide précieuse, tout particulièrement sur les questions d'insertion professionnelle. Comme pour les experts internationaux, leur degré de connaissance préalable du système universitaire est un élément important dans leur recrutement. Enfin, plus marginalement, certaines agences ont recours à des experts spécialistes de la qualité provenant d'autres agences d'évaluation.

L'élargissement de la composition du panel à des experts non universitaires pose des problèmes nouveaux que les agences doivent résoudre, notamment au regard de la légitimité d'une évaluation qui ne reposerait plus uniquement sur les pairs.

#### *L'usage de critères et de standards d'évaluation*

L'utilisation de critères et de références d'évaluation se fait généralement dans une approche adaptée aux objectifs poursuivis par les établissements ou par les gouvernements. Cette approche -en anglais *fitness for purpose* - vise à améliorer la qualité des programmes et des établissements d'enseignement supérieur dans leur contexte national ou local et en fonction des spécificités du système national d'enseignement supérieur et de ses besoins.

La distinction entre références et critères est difficile à établir. On pourrait assimiler la référence à un objectif à atteindre, sans nécessairement fixer un seuil qui soit mesurable. On pourrait appeler critère l'élément qui permet de justifier que la référence est atteinte ou en voie de l'être. Le critère serait l'élément d'appréciation, tandis que la référence correspondrait à une attente de la société, du gouvernement ou des partenaires à l'égard des universités. Mais références et critères ne sont jamais statiques et demeurent, quelque part, subjectifs.

Face à ces incertitudes, les agences ont souvent choisi de travailler plus directement avec les établissements et les conférences de présidents, voire avec les parties prenantes (employeurs ou fédérations étudiantes). Elles publient dorénavant leurs références et leurs critères d'évaluation. La publication des résultats d'évaluation demandée par les ministres européens est aussi un moyen d'inciter les agences à fonder leurs jugements sur des fondements plus précis et surtout plus explicites.

### **III - La convergence en matière d'évaluation de l'enseignement supérieur**

En Europe coexistent différentes conceptions sur l'utilité de l'évaluation de l'enseignement supérieur (Sursock, Bologna Handbook, 2006) : informer la société, inciter les établissements à s'améliorer, garantir l'existence de la qualité dans les formations proposées, réguler l'offre de formation nationale et européenne, protéger les consommateurs d'une offre de formation de mauvaise qualité, classer les meilleurs établissements... Le débat n'est pas tranché.

#### **Plusieurs mouvements de fond tendent à favoriser une certaine convergence**

##### *La structuration des acteurs et l'émergence des référentiels d'évaluation*

Les acteurs de l'enseignement supérieur et de l'évaluation se structurent. Ainsi, des organisations comme l'ENQA, l'EUA, l'ESIB, l'EURASHE sont devenues des acteurs clés (désignés par l'acronyme *E4*) pour les gouvernements et l'Union européenne. Elles facilitent les échanges de pratiques, conduisent des études transnationales, sont des forces de proposition.

Les référentiels d'évaluation se multiplient en Europe et dans le monde : les références et lignes directrices européennes en matière d'assurance de la qualité de l'enseignement supérieur, le *Guidelines of good practice* de l'*International Network for Quality Assurance Agencies in Higher Education* (INQAAHE), le *Code of good practice* de l'*European Consortium for Accreditation* (ECA), les lignes directrices pour la formation transnationale de l'OCDE-UNESCO, les codes d'accréditeurs professionnels... Les objectifs des textes sont évidemment différents, mais ils partagent un grand nombre de points communs (Aelterman, 2006).

En Europe, les références et lignes directrices européennes acceptées à Bergen en 2005 ont pour objectifs d'améliorer la qualité de l'enseignement supérieur, de

favoriser la culture de l'assurance qualité et d'informer les universités, les étudiants et la société sur les processus et les résultats. Elles visent également à contribuer à l'élaboration d'un cadre de référence commun au niveau de l'espace européen de l'enseignement supérieur. Elles ne se situent toutefois pas dans une perspective d'une reconnaissance européenne des diplômes.

Ont été définies sept références pour l'assurance de la qualité interne, qui constituent autant d'objectifs que les établissements doivent chercher à atteindre dans l'exercice de leurs missions, huit références pour l'évaluation externe, qui constituent les règles du jeu de l'évaluation que les agences d'évaluation doivent respecter et enfin huit références pour l'assurance qualité propres aux agences d'évaluation. L'évaluation de la qualité dans l'espace européen de l'enseignement supérieur dépend donc d'un savant équilibre entre les établissements évalués et les agences qui évaluent. Chacun est responsable de sa qualité et aucun n'est détenteur unique du contrôle de la qualité de l'enseignement supérieur. Les références européennes rendent explicites les engagements de chaque partie.

De grands principes cadrent les références et lignes directrices européennes. Ces principes respectent la diversité des systèmes d'enseignement supérieur, considérant que les activités des agences reflètent le contexte légal, social et culturel de leur pays et que leurs façons d'opérer dépendent d'objectifs assignés par le pouvoir politique national. Ils confirment par ailleurs que les références et lignes directrices européennes ne sont pas une liste de prescriptions détaillées à suivre impérativement et qu'elles ne conduisent en aucun cas à un système unifié européen d'assurance qualité.

*Des initiatives en faveur de la convergence dans l'enseignement supérieur*

De nombreuses initiatives sont prises par les universités, leurs réseaux, les États ou l'Union européenne pour favoriser la convergence des pratiques d'évaluation. On peut citer :

- le projet de l'EUA, *Quality Culture Project*, dont le but est de développer la culture qualité interne des établissements sans affaiblir leur diversité et d'instaurer des procédures d'évaluation externe de la qualité qui n'entravent pas l'activité des établissements ;
- les 10 règles d'or pour établir des masters conjoints et les lignes directrices pour l'amélioration de la qualité des programmes conjoints (projet *European Masters New Evaluation Methodology*, EUA, 2006) ;
- la *Joint Quality Initiative*, fondée en 2001, qui compte onze États membres (agences nationales d'accréditation ou représentants des ministres de l'Éducation). Ce réseau souhaite élaborer des descripteurs génériques de qualifications correspondant aux diplômes de licence, de master et de doctorat applicables à travers toute

l'Europe. Les descripteurs adoptés en 2002 et sont connus sous le nom de "descripteurs de Dublin" ;

- les études *Transnational European Evaluation Project 1* (TEEP 1, portant sur l'évaluation du 1<sup>er</sup> cycle de trois disciplines) et TEEP 2 portant sur l'évaluation des programmes ERASMUS MUNDUS, pilotées par ENQA,
- le projet *TUNING*, projet pilote lancé en 2000 par un groupe d'universités, visant à la convergence des structures éducatives en Europe à travers la définition des points de référence de compétences dans une série de disciplines.

#### *Une tentative de reconnaissance des décisions d'accréditation*

Plusieurs projets entendent aller plus loin dans l'harmonisation de l'enseignement supérieur au niveau européen, en favorisant la reconnaissance des diplômes nationaux et celle des décisions nationales d'accréditation.

La démarche conduite par l'*European Consortium for Accreditation* (ECA) est originale de ce point de vue. ECA est un réseau créé en novembre 2003 qui compte quinze agences d'accréditation. Son objectif est de parvenir à la reconnaissance mutuelle des décisions d'accréditation rendues par les participants avant la fin de l'année 2007. Les membres d'ECA estiment que cette reconnaissance mutuelle aidera à la reconnaissance des diplômes d'enseignement supérieur et à la mobilité des étudiants et des diplômés en Europe.

#### *Des études pour analyser les moyens de converger en évaluation en Europe*

Longtemps, les principales tentatives de convergence de l'assurance qualité en Europe ont été soit de compiler les références et les indicateurs utilisés dans plusieurs agences et présentant des caractéristiques communes, soit de rassembler des descriptions détaillées des méthodes d'évaluation ou d'accréditation appliquées au sein de chaque système national.

Cette compilation ne permet pas, en termes de pertinence et d'utilité, une analyse de l'application ou du transfert qui peut en être fait dans les systèmes nationaux. De même, la simple description des méthodes en usage au sein des agences ne peut fournir une information précise et actualisée sur les dynamiques internes et externes des systèmes nationaux d'enseignement supérieur. L'étude d'ENQA sur les possibilités de convergence, appelée *Quality convergence study* (ENQA, Crozier, Curvale, Hénard, 2005) a envisagé de substituer à la description des méthodes une meilleure compréhension des raisons pour lesquelles les méthodes étaient employées dans tel ou tel pays. Les principales conclusions de l'étude ont été les suivantes :

- l'objectif de convergence ne peut se satisfaire de la recherche d'une similitude de fonctionnement des systèmes d'enseignement supérieur ;

- une approche systémique met en évidence les interactions des différents acteurs et partenaires (enseignants, chercheurs, professionnels, étudiants, autorités, agences d'évaluation, etc.), chacun d'eux ne pouvant avoir qu'une vision partielle ou particulière du système ;
- l'approche systémique implique la reformulation de la question de la qualité. Indépendamment de ses définitions, la qualité est le produit d'une interaction entre les acteurs ou les éléments au sein du système. Elle est la conséquence de la façon dont le système fonctionne.

Pour un fonctionnement efficace des systèmes d'enseignement supérieur, il est impératif que tous les acteurs aient confiance en la façon dont les méthodes sont mises en œuvre au sein du système (par exemple, la politique d'admission des étudiants, l'organisation des examens, etc.). Mais la complexité de ces systèmes induit que la confiance ne peut simplement reposer sur la parfaite connaissance et la compréhension totale que l'on peut en avoir. La convergence et la confiance sont, en fait, intimement liées. La notion de convergence ne sous-entend pas une harmonisation des procédures et des méthodes d'évaluation qui convergeraient par la suite, mais elle renvoie plutôt à une application de références partagées par ces agences dans le but d'assurer leur indépendance, leur objectivité et de produire des jugements fondés sur des preuves (Surssock, 2006).

### **Quelles pistes pour favoriser la convergence en Europe ?**

Deux grand axes semblent se dessiner actuellement. L'un porterait sur le renforcement de la crédibilité des agences d'évaluation, l'autre viserait à favoriser la compréhension mutuelle entre les systèmes d'évaluation et d'assurance qualité nationaux.

#### *Améliorer la confiance dans l'évaluateur*

Puisque l'évaluation devient un outil stratégique censé *in fine* améliorer et/ou garantir la qualité de l'enseignement supérieur national et dorénavant européen, l'agence qui mène cette évaluation doit être crédible et ses jugements évaluatifs compréhensibles par tous.

L'adoption des références et lignes directrices européennes devrait fournir aux agences d'évaluation le cadre et les objectifs nécessaires pour améliorer les pratiques et les méthodes d'évaluation. Pourtant, leur mise en place opérationnelle pose des difficultés aux agences, notamment parce que leurs interprétations restent marquées par les considérations nationales.

À Bergen, les ministres ont demandé la mise en place d'un registre des agences dans le but d'y reconnaître les agences de qualité. Ce registre couvre plusieurs dimensions. Il peut être une liste des agences d'évaluation dont le travail répond à des critères de qualité et d'indépendance. Le registre peut aussi s'avérer un

outil de moralisation de l'offre d'accréditation, afin de lutter contre les accréditations frauduleuses. Dans cette perspective, le répertoire aurait pour fonction de repérer les agences offrant des garanties de sérieux, sans exclure pour autant les organismes d'accréditation à but lucratif qui s'implantent en Europe. Enfin, le registre peut faciliter la reconnaissance mutuelle des formations à l'échelle de l'Union européenne.

Dans la perspective du registre, plusieurs agences ont entamé leur évaluation, interne et externe qui serait une condition préalable à leur inscription et une condition d'adhésion à l'ENQA. Pour le moment, ces évaluations visent à montrer que les agences se préoccupent de leur qualité interne et qu'elles appliquent les références européennes.

#### *Travailler sur la compréhension mutuelle*

Pour parvenir à établir un cadre européen de l'évaluation et de l'assurance qualité, il serait sans doute nécessaire de travailler sur des concepts, un langage, des valeurs et des références communs et donc sur une culture européenne de la qualité des établissements (Williams, 2006). Cette démarche s'articulerait correctement avec les objectifs du processus de Bologne en matière d'assurance qualité. Elle ne devrait pas concerner uniquement les agences, mais favoriser les échanges entre tous les acteurs de l'enseignement supérieur, universités, agences, autorités de tutelles, étudiants, employeurs. D'autre part, les échanges devraient prendre une dimension sur le sens et les objectifs de l'évaluation et de l'assurance qualité.

Les chantiers en cours dans le milieu universitaire et celui de l'évaluation de l'enseignement supérieur montrent que ce mouvement est lancé. Le séminaire de l'ENQA de juin 2006 à Warwick, portant sur le langage de l'assurance qualité, le 1<sup>er</sup> forum européen pour l'assurance qualité réunissant les universités et le *E4* en novembre 2006 à Munich ainsi que les différents séminaires à venir sur la mise en place des références européennes en sont quelques récentes illustrations.

Malgré ces initiatives intéressantes, le sujet de l'évaluation de la qualité de l'enseignement supérieur demeure insuffisamment analysé et comparé dans une perspective historique, politique et épistémologique. Il semble aujourd'hui nécessaire d'engager des recherches dans ces directions.

**Références bibliographiques et sites Internet consultés**

**Aelterman, Guy (2006).** *Sets of standards for QA agencies: a comparison*. Den Haag : mis en ligne sur le site Internet du NVAO : [http://www.nvaio-event.net/inqaah/doc/Standards\\_QA\\_agencies-comparison.doc](http://www.nvaio-event.net/inqaah/doc/Standards_QA_agencies-comparison.doc)

**ANECA (2006).** *Implementing Quality Assurance standards in the European Higher Education Area. The way forward*. Compte rendu en ligne, site Internet de l'ANECA.

[http://www.aneca.es/actividades/eventos/activ\\_uimp06\\_eng.html](http://www.aneca.es/actividades/eventos/activ_uimp06_eng.html)

**Association des universités européennes (2006).** *Bologna Handbook, Making Bologna work*. Bruxelles : EUA.

<http://www.bologna-handbook.com>  
**Comité national d'évaluation (2004).** *Actes du colloque pour les 20 ans du CNÉ : de Berlin à Bergen - "Nouveaux enjeux de l'évaluation" (Dijon 10 et 11 juin 2004)*. Contient les définitions propres à l'assurance-qualité et à l'évaluation dans l'enseignement supérieur. Paris : mis en ligne sur le site Internet du CNÉ : [http://www.cne-evaluation.fr/WCNE\\_pdf/Colloquecne.pdf](http://www.cne-evaluation.fr/WCNE_pdf/Colloquecne.pdf)

**Commission européenne (2006).** *Recommandations de l'Union Européenne sur la qualité dans l'enseignement supérieur*. Bruxelles.

[http://www.europa.eu.int/eur-lex/lex/LexUriServ/site/fr/oj/2006/l\\_064/l\\_06420060304fr00600062.pdf](http://www.europa.eu.int/eur-lex/lex/LexUriServ/site/fr/oj/2006/l_064/l_06420060304fr00600062.pdf)

**Conférence des ministres européens (2005).** *Communiqué de la Conférence des Ministres européens chargés de l'Enseignement Supérieur, Bergen*.

[http://www.education.gouv.fr/sup/lmd/presente/docs/Bergen\\_communique\\_version\\_v\\_finaleFR.pdf](http://www.education.gouv.fr/sup/lmd/presente/docs/Bergen_communique_version_v_finaleFR.pdf)  
**Conseil de l'Europe (2006).** *La légitimité de l'assurance qualité dans l'enseignement supérieur*. Strasbourg : EUA.

[http://www.coe.int/t/dg4/highereducation/QA/Recommandations\\_FR.asp#TopOfPageContributions](http://www.coe.int/t/dg4/highereducation/QA/Recommandations_FR.asp#TopOfPageContributions)

**ENQA (2003).** *Enquête sur les procédures d'assurance-qualité en Europe. ENQA occasional paper 5*. Helsinki : Multiprint.

<http://www.enqa.eu/files/procedures.pdf>

**ENQA (2005).** *Atelier : Après la réunion de Bergen, résultats et inventaire sur la subsidiarité et la convergence*. Compte rendu en ligne, site Internet de l'ENQA.

<http://www.enqa.eu/eventitem.lasso?id=34851&cont=pasteventDetail>

**ENQA (2005).** *Atelier sur l'amélioration des méthodes d'évaluation*. Compte rendu en ligne, site Internet de l'ENQA.

<http://www.enqa.eu/eventitem.lasso?id=34868&cont=pasteventDetail>

**ENQA (2005).** *Quality Convergence Study 1. ENQA occasional paper 7*. Helsinki : Multiprint.

<http://www.enqa.eu/files/Quality%20Convergence%20Study.pdf>

**ENQA (2005), traduction CNÉ (2006).** *Références et lignes directrices pour le management de la qualité en Europe*. Helsinki : Multiprint.

[www.cne-evaluation.fr/WCNE\\_pdf/ESGCNE\\_vFrance.pdf](http://www.cne-evaluation.fr/WCNE_pdf/ESGCNE_vFrance.pdf)

**ENQA.** *Autres publications : TEEP 1, TEEP 2, Accreditation models in Europe, ...*) à consulter sur le site Internet de l'ENQA.

<http://www.enqa.eu/pubs.lasso>

**État de l'avancée du Processus de Bologne** - Bologna stocktaking.

<http://www.bologna-bergen2005.no/b/Stocktaking>

**Harvey, Lee (2004).** *War of the worlds: who wins in the battle for quality supremacy.* Quality in Higher Education, Volume 10, Number 1, éditeur <http://www.tandf.co.uk/journals/carfax/13538322.html>

**Harvey, Lee et Newton, Jethro (2004).** *Transforming quality evaluation.* Quality in Higher Education, Volume 10, Number 2 éditeur : <http://www.tandf.co.uk/journals/carfax/13538322.html>

**Haug, Guy (2003).** *Quality assurance / accreditation in the emerging European higher education area : a possible scenario for the future.* European Journal of Education.

<http://www.pfmb.uni-mb.si/bologna/quality%20assurance%20full%20txt.htm>

**Kahn, Sylvain (2003).** *L'espace européen d'enseignement supérieur à la croisée des chemins.* Revue en ligne "Études Européennes" <http://www.etudes-europeennes.fr>

**Office des publications des Communautés européennes (2006).** *Histoire de la coopération européenne dans le domaine de l'éducation et de la formation.*

(<http://publications.eu.int>)

**Perellon, J.-F. (2004).** *Approche comparée de politiques d'assurance-qualité des systèmes d'enseignement supérieur européens.* Observatoire science, politique et société, École polytechnique fédérale de Lausanne, Suisse.

<http://www.unige.ch/fapse/pegec/HES.pdf>

**Reichert, Sybille et Tauch, Christian (2005).** *Trends IV :European Universities Implementing Bologna.* Bruxelles. EUA.

[http://www.eua.be/eua/jsp/en/upload/TrendsIV\\_FINAL.1117012084971.pdf](http://www.eua.be/eua/jsp/en/upload/TrendsIV_FINAL.1117012084971.pdf)

**Van Damme, Dirk (2003).** *Standards and indicators in institutional and programme accreditation in higher education: A Conceptual Framework and a Proposal Paper for the Unesco-Cepes project.*

<http://www.cepes.ro/publications/pdf/Indicators.pdf>

**Williams, Peter (2006).** *Quality Assurance: the United Kingdom and Europe,* Sofia

<http://www.neaa.government.bg/images/stories/Peter%20Williams%20Bulgaria%2014%20Jun%2006.pdf>

**Sites Internet utiles****Classements d'universités :**

- Centre pour le développement de l'enseignement supérieur, Allemagne : <http://www.das-ranking.de/che7/CHE>

- Times Higher education, Royaume-Uni : <http://www.thes.co.uk/worldrankings/>

- Université de Shanghai : <http://ed.sjtu.edu.cn/rank/2005/ARWU2005Main.htm>

**Descripteurs de Dublin**

<http://www.jointquality.org>

**ECA (European Consortium for Accreditation)**

<http://www.eaconsortium.net/>

**ENQA** (ce site renvoie également à toutes les agences membres de l'ENQA)

<http://www.enqa.eu/>

**ESIB** - *The National Unions of Students in Europe*

<http://www.esib.org/>

**INQAAHE** *International Network for Quality Assurance Agencies in Higher Education*

<http://www.inqaahe.org/>

**EUA, 2003.** *Projet culture qualité interne.* Compte rendu en ligne, site Internet de L'EUA.

[http://www.eua.be/eua/en/projects\\_quality.jsp](http://www.eua.be/eua/en/projects_quality.jsp)

**EURASHE** : *European Association of Institutions in Higher Education* (secteur non universitaire)

<http://www.eurashe.be/>

**Évolution du système d'assurance-qualité au Royaume-Uni**

<http://www.aqa.ac.uk/publications/conferenceproceedings/2003derby/peterwilliams.htm>

**Projet TUNING**

[http://www.europa.eu.int/comm/education/policies/educ/tuning/tuning\\_fr.html](http://www.europa.eu.int/comm/education/policies/educ/tuning/tuning_fr.html)

**UN NOUVEAU PARADIGME POUR LES ENQUETES INTERNATIONALES:  
D'UNE PERSPECTIVE SYNCHRONIQUE VERS UNE PERSPECTIVE  
DIACHRONIQUE**

**Professeur Christian MONSEUR**, Université de Liège

Depuis le début des années nonante, en réponse à une demande de plus en plus forte des responsables des politiques éducatives, les surveys internationaux en sciences de l'éducation ont adjoint aux objectifs poursuivis traditionnellement la publication d'indicateurs de tendance, indicateurs destinés à suivre l'évolution des performances de populations scolaires. En témoigne le nom donné aux études contemporaines : Programme International de Suivi des Acquis des élèves (PISA), Progress International Reading Literacy Study (PIRLS), Trends International Mathematics and Science Study (TIMSS).

Alors qu'une différence de 10 points sur une échelle dont l'écart-type est généralement fixé par convention à 100 ne change pas l'interprétation pédagogique des résultats pour une collecte de données, une telle différence prend une toute autre dimension pédagogique et politique si elle représente le changement de performance entre deux collectes de données.

Pour être valable, cette comparaison diachronique suppose que la méthodologie, au sens large du terme, soit parfaitement identique pour les deux collectes de données, ou que les éventuels changements méthodologiques n'affectent en rien la comparabilité temporelle.

A travers quelques exemples issus des données PISA pour l'essentiel, ce pré-supposé sera confronté à la réalité empirique ; les résultats montrent la nécessité d'une grande prudence à l'égard de ces indicateurs de tendance.