

L'évaluation des acquis des étudiants dans l'enseignement universitaire

Marc ROMAINVILLE

Professeur au département Education et Technologie
des Facultés universitaires de Namur (Belgique)

Décembre 2002

Rapports établis à la demande du Haut Conseil de l'évaluation de l'école

Directeur de la publication : Christian FORESTIER

Secrétariat général : 3-5, bd Pasteur 75015 – PARIS

Tel : 01 55 55 77 14

Mèl : hcee@education.gouv.fr

ISSN en cours

Conception et impression : DEP/BED

L'ÉVALUATION DES ACQUIS DES ÉTUDIANTS DANS L'ENSEIGNEMENT UNIVERSITAIRE

1 INTRODUCTION

1.1 OBJECTIFS ET PLAN DU RAPPORT

L'objectif général du présent rapport est de dresser un état des connaissances dont on dispose à propos de l'évaluation des acquis cognitifs des étudiants. Le rapport est articulé autour des quatre questions qui suivent.

1.1.1 Que sait-on des acquis des étudiants ?

La première étape de la mission a consisté à rassembler les informations disponibles sur les acquis des étudiants. Que sait-on de leurs acquis aux principaux paliers de la formation tertiaire, soit aux différentes portes de sortie qui s'offrent à eux ? En d'autres termes, de quelles informations dispose-t-on à propos des connaissances et des compétences acquises par les étudiants au cours de leurs études ? Que savent-ils et que savent-ils faire ?

Il est vite apparu que très peu de données étaient disponibles sur le sujet, que ce soit au plan local (à l'échelle d'un établissement, d'un groupe d'établissements ou d'une académie) ou au plan national. Au plan international, les grandes opérations de mesure des acquis telles qu'elles sont, par exemple, organisées régulièrement par l'OCDE ou l'IEA s'arrêtent aux portes de la scolarité obligatoire et ne concernent généralement pas le supérieur. Après avoir résumé les quelques rares données collectées (comme les études américaines sur les effets de l'enseignement supérieur), le travail s'est ensuite orienté vers des tentatives d'explication du peu de données disponibles, en cherchant à décrire les pratiques d'évaluation des acquis qui ont cours à l'université.

1.1.2 Pourquoi sait-on si peu de choses sur les acquis des étudiants ? Les pratiques d'évaluation à l'université

Face au déficit de données précises sur les acquis des étudiants, l'étape suivante de la mission a consisté à s'interroger sur les pratiques d'évaluation, en faisant l'hypothèse que l'étude de ces pratiques permettrait de mieux comprendre le peu de données standardisées qu'elles produisent. On s'est ainsi interrogé sur les caractéristiques des examens tels qu'ils sont organisés dans l'enseignement universitaire. Que sait-on des pratiques d'évaluation des acquis des étudiants ?

Bien qu'elles occupent une place importante dans l'exercice du métier d'enseignant-chercheur et qu'elles conditionnent dans une large mesure l'apprentissage des étudiants, les pratiques d'évaluation restent mal connues. L'hétérogénéité est sans doute leur premier trait dominant : on observe en effet une absence de standardisation des dispositifs, des procédures, des exigences et des critères sur la base desquels les acquis des étudiants sont appréciés. Cette importante hétérogénéité des pratiques nuit à la fidélité et à la validité de l'évaluation. On a aussi cherché à dresser un tableau différencié de cet état de fait, en tenant compte des caractéristiques des filières qui peuvent l'influencer. Par exemple, l'étude de cas réalisée dans le cadre de la présente mission montre que le concours d'agrégation engendre un effet régulateur sur la standardisation des objectifs et des contenus, largement en amont de cette épreuve.

Le rapport se penche aussi sur les nombreux facteurs qui concourent à cette hétérogénéité des pratiques d'évaluation : la diversité des objectifs et des contenus des cursus, même pour un diplôme national ; la longue tradition « humboldtienne » de liberté académique, qui reste tenace à l'Université ; le fait que l'évaluation des acquis cherche à remplir des fonctions très différentes, voire inconciliables.

Le second trait dominant des pratiques d'évaluation des acquis réside dans la faible explicitation dont elles font l'objet. Autrement dit, non seulement les étudiants ne sont pas évalués de la même manière, mais on ne sait pas non plus très bien sur quels critères ils sont évalués. L'évaluation ne se réalise en effet pas, de manière prioritaire, en regard d'inventaires explicites des connaissances et des compétences qui sont attendues en fin de formation. Le rapport cherche aussi à mettre en lumière les facteurs qui sont susceptibles d'expliquer cette absence d'explicitation des critères à partir desquels l'évaluation est menée et le développement corrélatif d'une évaluation normative.

1.1.3 Pourquoi devrait-on en savoir davantage ? Faiblesses et lacunes de l'évaluation des acquis

L'absence de données, de standardisation et d'explicitation étant plutôt la règle que l'exception, l'étape suivante de la mission a consisté à s'interroger sur les raisons que l'on aurait à prôner une modification de cet état de fait qui perdure dans le supérieur depuis de nombreuses années. La piste privilégiée a été de dresser un inventaire des effets dommageables de l'absence de données sur les acquis et, plus généralement, des faiblesses et des lacunes des pratiques actuelles d'évaluation à l'université. Ces effets négatifs définissent, en creux, les raisons au nom desquelles devraient être entreprises des actions visant à mieux connaître les acquis des étudiants et à améliorer la qualité des pratiques d'évaluation.

1.1.4 Comment améliorer les pratiques d'évaluation des acquis ?

Quels dispositifs s'agirait-il de mettre en place de manière à obtenir des informations sur les acquis des étudiants et, plus généralement, de manière à améliorer la qualité des pratiques d'évaluation ? Cette dernière partie du rapport vise à suggérer quelques pistes d'action à entreprendre pour pallier les faiblesses de l'évaluation actuelle des acquis et pour atténuer ses effets dommageables.

1.2 MÉTHODOLOGIE

Il est vite apparu que le traitement de la question posée par le Haut Conseil ne pouvait pas se restreindre au seul recensement des études sur les acquis des étudiants, sous peine de tourner vite court. Par ailleurs, l'exploration de cette question aboutit inéluctablement à reposer un certain nombre de questions fondamentales quant au fonctionnement général des formations universitaires : que valident les diplômes ? Quelle est la place des examens dans le parcours de formation ? Quelles sont les finalités de l'enseignement tertiaire ? Les compétences mesurées ont-elles bien fait l'objet d'un apprentissage systématique préalable ? L'évaluation dans le supérieur cherche-t-elle à dresser un bilan des acquis ou à classer les étudiants ? ...

La méthodologie mise en œuvre pour réaliser cette mission a cherché à tenir compte de cette double spécificité, en multipliant et en diversifiant les sources d'information. Trois méthodes complémentaires ont été privilégiées.

- Premièrement, l'analyse de la situation actuelle, articulée autour des quatre questions évoquées ci-dessus, a pris appui sur une **étude de documents** : littérature pédagogique francophone et anglophone ; actes de colloques d'associations de pédagogie dans le supérieur (Association internationale de pédagogie universitaire, Association pour le développement des méthodes dans l'enseignement supérieur, Society for Research in Higher Education) ; revues de pédagogie dans

l'enseignement supérieur¹ ; rapports prospectifs sur l'enseignement supérieur ; rapports du CNE, de la CPU, de l'AMU, ...

- Des **entretiens semi-dirigés** ont aussi été menés, autour des quatre mêmes questions, auprès d'un nombre limité de responsables pédagogiques (président du CNE, responsable des DEUG à la direction de l'enseignement supérieur, directeurs d'UFR et enseignants-chercheurs).
- Une troisième piste a consisté à procéder à une **étude de cas**, visant à comparer, de manière qualitative et dans le détail, les pratiques d'évaluation des acquis des étudiants d'une même filière dans deux universités différentes : quelles sont les modalités d'examen ? Quels sont les critères d'évaluation ? Quelles sont les compétences qui sont mesurées sur les deux sites ? Qu'ont-elles en commun ? La diversité des pratiques d'évaluation est-elle le reflet de différences d'objectifs ? Cette étude de cas, qui a contribué à enrichir l'ensemble du rapport, est présentée en annexe 1 (7.1).

Le bilan sur l'évaluation des acquis des étudiants, tel qu'il est présenté dans ce rapport, revêt un caractère exploratoire évident. Les données sur les acquis cognitifs restent éparpillées, les inventaires de connaissances et de compétences rares. Les recueils d'informations qui ont été réalisés pour nourrir le rapport représentent des coups de sonde qualitatifs : une étude de cas et une dizaine d'entretiens. Les études sur les pratiques d'évaluation se sont révélées plus nombreuses, mais il reste que le bilan général dressé dans ce rapport mériterait d'être élargi, affiné et davantage documenté, notamment par l'enregistrement plus systématique des « bonnes pratiques » d'évaluation des acquis des étudiants dans l'enseignement universitaire.

¹ Les numéros des dix dernières années des revues suivantes ont été consultés : Higher Education, Studies in Higher Education, Teaching in Higher Education, Research in Higher Education abstracts, Assessment and Evaluation in Higher Education, Res Academica et la revue « Gestion de l'enseignement supérieur » de l'OCDE.

2 QUE SAIT-ON DES ACQUIS DES ÉTUDIANTS ?

De manière générale, la question de l'évaluation des acquis des étudiants occupe une place relativement minime dans la littérature pédagogique consacrée à l'enseignement supérieur². En particulier, la littérature fournit très peu de données quantifiées sur les acquis des étudiants du supérieur.

À notre connaissance, aucune enquête n'a cherché à identifier et à mesurer, à l'échelle d'un système éducatif, les acquis des étudiants aux différents paliers de leurs études. De la même manière, aucun des programmes internationaux de mesure des acquis, comme ceux qui sont régulièrement menés par l'OCDE (par exemple le récent programme PISA) ou par l'IEA, n'a spécifiquement porté sur ce niveau des systèmes éducatifs. Il semble, mais le caractère exhaustif de l'inventaire est ici beaucoup plus sujet à caution, qu'il en va de même pour les enquêtes plus locales, au niveau d'une région, d'une académie, voire d'un établissement ou d'un groupe d'établissements.

Les rares études qui fournissent quelques indications sur les acquis des étudiants se regroupent autour de quatre thèmes : les compétences des étudiants à l'entrée des études, leur perception des compétences acquises à la sortie, leurs compétences méthodologiques et les effets généraux de l'enseignement supérieur.

2.1 DES INVENTAIRES DE COMPÉTENCES D'ENTRÉE

En amont de la formation universitaire, des expériences de mesure des compétences des étudiants à l'entrée des études sont rapportées ci et là. L'expérience belge MOHICAN, par exemple, a consisté à mesurer, dans l'ensemble des universités de la partie francophone du pays, la maîtrise d'un certain nombre de compétences estimées être des prérequis aux différentes filières universitaires³. Des expériences de mesure plus locale, à l'échelle d'une institution ou d'une filière, sont également rapportées ponctuellement. Par exemple, Lebrun et Lega (1999) ont élaboré un test portant sur la maîtrise de compétences scientifiques transversales (ex. savoir lire et interpréter un graphique). Ce test permet de dresser un portrait des préacquis généraux des étudiants à l'entrée des filières scientifiques. Des tests de même nature ont aussi été développés dans le domaine de la maîtrise de la langue maternelle (Defays et al., 2000).

2.2 DES ÉTUDES SUR LA PERCEPTION QU'ONT LES ÉTUDIANTS ET LEURS EMPLOYEURS DES COMPÉTENCES ACQUISES

En aval de la formation, des études ont également été consacrées à la perception qu'ont les **étudiants** des compétences qu'ils estiment avoir développées au cours de leurs études supérieures. Par exemple, Drew (1998) a mené, par entretiens de groupe puis par questionnaires à grande échelle, une étude sur les compétences que les étudiants pensent avoir acquises durant leurs études supérieures.

² Il s'agit même d'un des points les plus obscurs de la pédagogie universitaire, qui pourtant n'en manque pas... En regard de l'importance grandissante que cette tâche prend désormais dans le travail de l'enseignant-chercheur (cf. 3.1) et compte tenu des nombreuses difficultés docimologiques qui y sont liées (cf.3.4), on ne peut qu'être surpris par le faible traitement de cette question dans la littérature. Ainsi, sur les dix dernières années, la principale revue internationale consacrée à la gestion de l'enseignement supérieur, « Higher Education », ne consacre pas plus de quatre articles à cette question. Plus étonnant, la revue « Teaching in Higher Education », entièrement dédiée à des questions pédagogiques, n'en compte qu'à peine plus du triple. Et même dans la revue « Assessment and Evaluation in Higher Education », la majorité des articles concernent soit l'évaluation des enseignements, soit les processus d'évaluation de la qualité, surtout des établissements. Il faut cependant noter que le troisième thème majeur de cette revue a trait aux pratiques innovantes d'évaluation des acquis des étudiants (cf. 5.2.4).

³ Une présentation succincte de cette expérience est disponible sur le site : <http://www.cfwb.be/ciuf/coordination/mohican.htm>

Dans le contexte d'un enseignement supérieur qui est désormais invité à produire des étudiants « employables » (cf. 4.1.2), le souci de cette étude a été de dresser un inventaire subjectif des compétences que ce niveau du système éducatif permet d'acquérir, en particulier des compétences génériques tant mises en avant dans les exigences actuelles des employeurs : adaptabilité, créativité, expression orale et écrite, capacité à travailler en groupe, ... L'annexe 7.2 synthétise les résultats de cette vaste enquête auprès des étudiants. On y découvre, de manière intéressante, que ce qui est directement mesuré par les examens universitaires traditionnels (la maîtrise de connaissances disciplinaires) n'est pas considéré, par les étudiants, comme le bénéfice principal de leurs études : ils ne citent, en effet, cet élément qu'en cinquième position.

Il existe aussi des études similaires consacrées à la perception qu'ont les **employeurs** des acquis des étudiants, mais cette évaluation externe sort du cadre de notre mission. La plupart du temps, il s'agit d'enquêtes qui portent sur la satisfaction (et plus souvent encore, sur l'insatisfaction...) exprimée par les employeurs par rapport aux compétences qu'ils estiment que les étudiants ont acquises durant leurs études supérieures (Hart, Bowden & Watters, 1999).

Il faut enfin signaler que nous disposons aussi d'études qui dressent l'inventaire des compétences telles qu'elles sont attendues par le monde du travail. Il ne s'agit évidemment plus, à proprement parler, de recensements d'acquis, même perçus, mais plutôt de souhaits, d'attentes, voire de doléances. Cependant, ces études apportent un éclairage intéressant à la discussion sur les finalités des formations universitaires : elles montrent notamment qu'il peut exister un décalage important entre les finalités académiques, souvent exprimées selon des logiques de contenus disciplinaires, et les attentes du monde économique et social. À titre d'exemple, un récent rapport du CNE (2002a) sur les formations supérieures en mathématiques orientées vers les applications contient une description intéressante des qualités attendues d'un cadre recruté à la sortie de ces filières. Cette liste de compétences a été élaborée à l'aide d'entretiens menés auprès d'un échantillon de responsables d'entreprises industrielles et de service qui embauchent régulièrement ce type d'étudiants. Comme nous l'avons signalé ci-dessus pour le point de vue des étudiants, ce sont surtout des compétences génériques qui sont mises en avant par ces employeurs (cf. tableau 1).

Tableau 1

Qualités attendues d'un cadre recruté dans le domaine des applications mathématiques
(d'après CNE, 2002a)

<p>1. Qualités intellectuelles et professionnelles</p> <ul style="list-style-type: none"> ▪ Capacité au raisonnement, à l'abstraction, à la conceptualisation et à la rigueur (ex. mobiliser son savoir pour extraire l'outil adéquat) ▪ Créativité (ex. créer de nouveaux outils) ▪ Connaissance des organisations (ex. anticiper les conséquences organisationnelles de l'introduction d'un nouvel outil) ▪ Maîtrise de l'incertitude et de la complexité ▪ Approche interdisciplinaire <p>2. Qualités personnelles et comportementales</p> <ul style="list-style-type: none"> ▪ Aptitude à l'écoute ▪ Attitude ouverte, curiosité ▪ Capacité à communiquer et à convaincre ▪ Compétences relationnelles ▪ Capacité à travailler en équipe sur des projets ▪ Capacité à « parler le langage de tous les savoirs »
--

2.3 DES INVENTAIRES DE COMPÉTENCES MÉTHODOLOGIQUES

On trouve aussi, dans la littérature, quelques enquêtes qui visent à dresser l'inventaire des **compétences méthodologiques des étudiants**. Ces études sont réalisées dans une perspective correctrice, en réponse à l'échec massif qui règne au premier cycle. Elles se fondent sur l'idée qu'un

problème majeur de la transition entre l'enseignement secondaire et l'enseignement supérieur réside dans la maîtrise progressive, par le jeune issu du secondaire, de son nouveau métier d'étudiant, ce nouveau métier impliquant notamment une série d'ajustements méthodologiques importants (Coulon, 1997). Les travaux de Boulet et al. (1996) sont représentatifs de ce courant de recherche. Une vaste enquête leur a permis d'identifier les compétences méthodologiques qui caractérisent l'étudiant qui réussit ses études universitaires, au sein des différents champs disciplinaires.

2.4 DES ÉTUDES SUR L'EFFET GÉNÉRAL DE L'ENSEIGNEMENT SUPÉRIEUR

Il faut aussi signaler que des études effectuées en Amérique du Nord ont porté sur **l'effet général de l'enseignement supérieur**. La synthèse la plus récente de ces nombreux travaux est celle qu'ont publiée, en 1991, Pascarella et Terenzini, sous le titre « *How college affects students* ». Cette synthèse ambitieuse vise à dresser un inventaire de l'impact des quatre années du « bachelor's degree » en termes d'acquis des étudiants. Elle est ambitieuse à plus d'un titre. D'abord, elle se fonde sur plus de 2.600 recherches qui ont été consacrées à cette question⁴. Ensuite, l'étude cherche à identifier les effets de l'enseignement tertiaire américain sous des angles très divers : les acquis cognitifs bien sûr (des connaissances disciplinaires au développement cognitif), mais aussi les modifications d'attitude, de valeurs et de personnalité. Elle s'intéresse également aux effets de la formation tertiaire sur la carrière, le niveau de vie et le salaire des étudiants. Enfin, elle vise à isoler ce que les auteurs appellent « l'effet net » de l'enseignement supérieur, c'est-à-dire les changements effectivement attribuables au passage de l'étudiant par l'enseignement supérieur, indépendamment des effets de la sélection et de la maturation. Pour ce faire, la synthèse se fonde en partie sur des études, trop rares cependant, qui ont comparé le développement de compétences des jeunes issus du bachelor's degree avec celles de jeunes du même âge, mais qui n'ont pas suivi cet enseignement tertiaire. Les effets sont très souvent mesurés à l'aide de tests standardisés, relativement prolifiques aux États-Unis, y compris dans le domaine des connaissances spécifiques. Par exemple, les « gains » des étudiants dans le domaine de la pensée critique sont établis sur la base du test « Watson-Glaser Critical Thinking Appraisal ». On tentera ci-dessous de résumer les principaux enseignements de la synthèse de Pascarella et Terenzini, en distinguant trois catégories d'effets : les acquis de connaissances et de compétences, les acquis de savoir-faire cognitifs généraux et les acquis non cognitifs.

2.4.1 Les acquis de connaissances et de compétences

D'après les auteurs, l'enseignement universitaire américain atteindrait assez bien ses objectifs en ce qui concerne l'acquisition, par les étudiants, de connaissances dans les matières liées à leur sujet principal d'études, ainsi qu'en ce qui concerne les compétences langagières et, dans une moindre mesure, mathématiques (Tableau 2).

⁴ Le lecteur pourrait trouver quelque peu contradictoire que nous ayons annoncé que les inventaires d'acquis étaient rares alors que cette métaanalyse porte sur plus de 2600 recherches... En fait, comme il s'agit d'un agrégat d'études ponctuelles, la métaanalyse ne fournit pas de données représentatives sur le type d'acquis des étudiants dans chaque discipline. La synthèse de Pascarella et Terenzini se fonde en effet sur une accumulation de recherches en éducation, très locales, réalisées sur des échantillons restreints et portant sur des progrès moyens, enregistrés à divers tests standardisés. Ces études n'offrent donc pas de données représentatives nationales sur les connaissances et les compétences précises que les jeunes Américains maîtrisent, dans les différentes disciplines, à la sortie de leurs études.

Tableau 2
Acquis cognitifs estimés de connaissances et de compétences
(d'après Pascarella & Terenzini, 1991)

Acquis	Ampleur estimée ⁵
Compétences langagières	21
Compétences mathématiques	10
Connaissances disciplinaires	31

L'essentiel des acquis semble être le fait des deux premières années d'études sur les quatre que compte le « bachelor's degree ». Les établissements qui assurent davantage d'acquis à leurs étudiants répondent aux caractéristiques suivantes : ils favorisent des interactions fréquentes entre les professeurs et les étudiants, leurs cursus sont flexibles et enfin leur personnel enseignant possède un niveau de qualification élevé. Au sein d'un même établissement, l'ampleur des acquis dépend de l'investissement de l'étudiant dans ses études et de certaines dimensions pédagogiques (la présence d'évaluations formatives, par exemple), mais pas d'autres : en particulier, les méthodes d'enseignement en petits groupes ne semblent pas systématiquement plus efficaces que l'enseignement magistral en grand groupe. Quant aux caractéristiques des enseignants qui assurent des acquis plus importants chez leurs étudiants, il est intéressant de constater qu'elles correspondent, dans une large mesure, aux critères que les étudiants retiennent quand il s'agit de décrire les enseignants qu'ils jugent « efficaces » (Romainville, 2000) : maîtrise de la matière, enthousiasme, clarté dans les explications, gestion efficace du temps, capacité de faire comprendre par des exemples et des analogies, langage clair et précis, capacité d'écoute des étudiants et de leurs problèmes.

Comme le notent eux-mêmes les auteurs de l'étude, il aurait été surprenant que ces acquis cognitifs de connaissances et de compétences n'aient pas eu lieu... Par contre, il est intéressant d'observer que les étudiants progressent aussi dans le domaine des savoir-faire cognitifs plus larges, comme l'indique le point suivant.

2.4.2 Les acquis de savoir-faire cognitifs et de développement intellectuel

Non seulement, les étudiants en savent plus à la sortie, mais plusieurs études montrent qu'ils ont aussi acquis une série de savoir-faire cognitifs qui leur permettront de continuer à apprendre tout au long de leur vie (par exemple, parce qu'ils auront développé des pratiques de lecture) et à traiter des problèmes nouveaux et complexes. Les principaux acquis répertoriés par la synthèse de Pascarella et Terenzini concernent, par ordre d'importance, les domaines suivants :

Tableau 3
Acquis de savoir-faire cognitifs estimés
(d'après Pascarella & Terenzini, 1991)

Acquis	Ampleur estimée
Capacité à traiter des problèmes complexes	38
Capacité à raisonner et à argumenter face à des problèmes ouverts	34
Pensée critique	34
Compétences communicationnelles à l'oral	22
Compétences communicationnelles à l'écrit	19

⁵ L'ampleur de l'acquis est exprimée en termes de gain de centiles, en fixant la moyenne des notes des étudiants à leur entrée au 50^e centile sur cent. Le tableau peut alors s'interpréter de la manière suivante : dans le domaine des compétences langagières, 21 centiles séparent l'étudiant moyen à son entrée de l'étudiant moyen à la sortie, c'est-à-dire après avoir bénéficié de la formation tertiaire.

Une rapide comparaison des tableaux 2 et 3 montre que les acquis de savoir-faire cognitifs semblent d'une plus grande ampleur que les acquis de connaissances, ce qui rejoint d'une certaine manière le point de vue des étudiants (cf. 2.2). Par exemple, en se basant sur les résultats d'élèves aux tests de raisonnement abstrait à la sortie du secondaire, Pascarella et Terenzini estiment que 50% des compétences de ce type, telles qu'on les observe chez les étudiants à la fin du « bachelor's degree », sont effectivement mises en place durant ces études. L'analyse des « effets nets » montre, de plus, que ces acquis sont principalement dus à l'enseignement supérieur lui-même et que l'essentiel des progrès enregistrés dans ces domaines lui est donc imputable. Les établissements les plus efficaces à développer des savoir-faire cognitifs généraux répondent aux caractéristiques suivantes : ils sont sélectifs, de petite taille et animés d'un « éthos institutionnel », comme l'appellent les auteurs, qui suscite un investissement de la part de l'étudiant. Les méthodes d'enseignement n'ont que peu d'influence sur ces acquis, à l'exception de l'ancrage de l'enseignement de ces savoir-faire au sein même des disciplines constitutives de la filière d'études. Autrement dit, le raisonnement critique a d'autant plus de chance d'être maîtrisé par les étudiants qu'il leur aura été enseigné à propos de faits liés à leur domaine d'études.

2.4.3 Autres acquis

La synthèse de Pascarella et Terenzini détaille enfin d'autres types d'effets, mais dont l'ampleur semble moindre et que nous ne ferons qu'évoquer dans le présent rapport consacré aux acquis cognitifs.

- **Changements d'attitudes et de valeurs** : intérêt grandissant pour les questions intellectuelles, culturelles et artistiques ; libération de l'individu par rapport aux influences, notamment religieuses ; développement d'une motivation éducative intrinsèque (défi intellectuel, compréhension du monde, ...) et diminution corrélative de la motivation extrinsèque et instrumentale (diplôme, sécurité d'emploi, ...).
- **Changements psychosociaux** : affirmation de soi, développement de l'indépendance, de la tolérance et de la maturité dans les relations interpersonnelles, même s'il est difficile de savoir si ces évolutions sont imputables à la formation elle-même. Par contre, la diminution du dogmatisme et de l'ethnocentrisme semble être un effet direct des études supérieures.
- **Effets sur les carrières** : le passage par l'enseignement supérieur assure l'accès à des carrières valorisantes et l'obtention d'un salaire élevé. Il augmente aussi la probabilité de trouver un emploi et garantit une certaine satisfaction au travail. Enfin, il a des effets positifs sur le niveau de vie, en ce compris la santé. Ces effets sont d'autant plus forts que les études sont longues et sélectives, ils s'observent dans certaines filières plus que dans d'autres (ingénierie, économie, ...).

Un autre résultat intéressant de cette synthèse est qu'on observe en définitive moins de différences entre les acquis d'étudiants provenant d'établissements différents (élitistes ou non, réputés ou non, exigeant des frais de scolarité importants ou non) qu'entre les étudiants qui font des études supérieures et ceux qui n'en font pas. Ce qui importe est donc de pouvoir bénéficier d'une formation tertiaire, davantage que le fait d'en bénéficier ici ou là. Il est d'ailleurs significatif de constater que la qualité de l'établissement (en termes de réputation, de taille et de sélectivité) a peu d'impact sur la qualité des acquis cognitifs alors que les effets socioéconomiques sur les carrières sont fortement tributaires du type d'établissement fréquenté. Autrement dit, le fait d'accéder à des institutions réputées n'assure pas nécessairement des acquis cognitifs nettement supérieurs, mais laisse augurer, sans doute par l'effet « signal » de ce parcours, d'un meilleur avenir socioprofessionnel.

Ces études sur l'impact du collège ont été produites en réponse aux critiques qui se faisaient jour aux États-Unis à propos de la qualité de la formation du premier cycle universitaire en regard de son coût grandissant. Elles présentent donc un caractère d'auto-justification assez évident. Il faut dès lors

indiquer qu'à côté de ces synthèses de recherches aux cadres psychologiques dominants⁶, des études plus critiques sur les effets de l'enseignement universitaire de masse ont vu le jour. Beattie (1995), par exemple, fait remarquer qu'il existe un écart considérable entre les objectifs prestigieux, de haut niveau, tels qu'ils ont été, de tout temps, présentés pour l'enseignement universitaire et la réalité des acquis dans un enseignement massifié et désargenté. Ainsi, la « formation de l'esprit » a toujours constitué une des finalités majeures de l'enseignement universitaire⁷, mais on peut se demander s'il ne s'agit pas d'un mythe, resté très vivace, en dépit des pratiques pédagogiques observées dans les grands groupes du premier cycle.

Ce type d'interrogation critique sur les acquis s'observe également dans de nombreux autres rapports, notamment ceux du CNE (cf. 3.5). Le récent rapport Lyon-Caen sur les études de droit (2002) dénonce aussi la difficulté que rencontrerait l'université à atteindre ses hautes ambitions de formation : *« l'accent est mis sur l'acquisition d'un savoir technique beaucoup plus que sur la maîtrise d'outils de raisonnement. Le juriste formé de la sorte s'apparente trop souvent à un ingénieur technologue, doté d'une connaissance précise de règles régissant tel ou tel domaine mais peu à même d'en relativiser le contenu, d'en saisir les enjeux, et de faire usage d'une véritable aptitude à la décision et à l'argumentation »* (Lyon-Caen, 2002, p.12).

Ces interrogations critiques ont reçu l'appui des études empiriques consacrées aux effets des pratiques d'évaluation sur les apprentissages des étudiants, dont nous parlerons au point 3.5.

Si, en conclusion de cette première partie, nous pouvons affirmer que l'on ne sait pas précisément ce que les étudiants ont acquis durant leurs études universitaires, rien ne nous permet d'en déduire, bien sûr et fort heureusement, qu'ils n'ont rien acquis. La relativement bonne « employabilité » des diplômés universitaires atteste sans doute d'acquis appréciés par le monde du travail. Les étudiants eux-mêmes évoquent d'importants bénéfices de leur passage par l'université. Les études sur les effets généraux de l'enseignement supérieur font également état de gains substantiels dans certains domaines. La conclusion principale, à ce stade du rapport, est triple :

- les acquis des étudiants ne sont pas suffisamment explicités, identifiés et répertoriés ;
- les acquis majeurs, aux yeux des étudiants et du monde du travail, ne sont peut-être pas ceux qui sont au cœur des examens traditionnels ;
- si l'université recentrait ses pratiques d'évaluation sur ces acquis essentiels, on peut faire l'hypothèse qu'elle parviendrait à les développer davantage.

⁶ Un grand nombre d'entre elles sont menées au sein de cadres théoriques issus de la psychologie du développement du jeune adulte.

⁷ « Former l'esprit en lui apprenant à bien raisonner sur toute chose, à chercher la vérité et à la saisir », tel était, au milieu du dix-neuvième siècle, la raison d'être de l'enseignement supérieur (Newman cité par Beattie, 1995).

3 POURQUOI SAIT-ON SI PEU DE CHOSES ? LES PRATIQUES D'ÉVALUATION DES ACQUIS À L'UNIVERSITÉ

Face au déficit de données précises sur les acquis, l'étape suivante a consisté à s'interroger sur les pratiques d'évaluation des étudiants, en faisant l'hypothèse que l'étude de ces pratiques permettrait de mieux comprendre le peu de données standardisées qu'elles produisent. Dans ce domaine aussi, le relevé de la littérature a vite montré que les études qui livrent des informations sur les pratiques d'évaluation dans le supérieur étaient peu nombreuses, sans être toutefois aussi rares que celles qui inventorient les acquis des étudiants.

Les **pratiques réelles d'évaluation** des acquis des étudiants sont **encore mal connues**. Par exemple, l'inventaire pourtant minutieux des tâches du professeur d'université réalisé par Bertrand et ses collègues (1994) ne fournit pas d'indications précises sur la part du travail professoral consacrée aux tâches d'évaluation et sur les comportements qu'elles impliquent. Ces tâches sont intégrées, et donc noyées en ce qui concerne les résultats, au sein de la rubrique générale « enseignement ». La même étude signale pourtant que l'évaluation des apprentissages est une composante importante des activités d'enseignement du professeur d'université (cf. Tableau 4), mais on sait peu de choses sur la manière dont il s'acquitte de cette mission.

Tableau 4

Activités constitutives de la tâche d'enseignement (d'après Bertrand et al., 1994, p. 131)

Activités d'enseignement	N	Nombre moyen d'heures par semaine	%
Prestation directe	2138	5,4	34,0
Préparation d'un cours	2137	4,7	29,4
Encadrement des étudiants	2136	2,6	14,2
Évaluation des apprentissages	2136	2,2	13,6
Gestion courante	2129	1,4	6,5

De même, on connaît mal la **perception qu'ont les étudiants des pratiques d'évaluation**. On sait quand même qu'une plus grande cohérence des pratiques d'évaluation et de notation constitue une de leurs principales revendications quand ils sont interrogés à propos des lacunes de l'enseignement universitaire (Blais et al., 1997 ; Gibbs, 1995). De même, quel que soit leur profil d'apprentissage, ils se montrent fort critiques sur les examens traditionnels. Ces derniers font, à leurs yeux, trop souvent appel à la seule mémorisation, pas assez à l'intégration des connaissances et les incitent à développer des apprentissages superficiels (cf. 3.5), qui ne leur seront d'aucune utilité dans leur vie professionnelle (Blais et al., 1997 ; LindBlom-Ylänne & Lonka, 2001). Cependant, ils estiment aussi que les pratiques innovantes d'évaluation, comme les travaux de groupes évalués par les pairs, sont moins équitables que les examens traditionnels (Norton & Brunas-Wagstaff, 2000). Par ailleurs, il nous a aussi été rapporté, au cours des entretiens, que les épreuves telles qu'elles sont organisées au sein des grands groupes du premier cycle ne semblent pas toujours avoir une grande crédibilité aux yeux des étudiants : ces derniers n'hésitent pas à mettre en doute la fiabilité des modes actuels d'évaluation des acquis (cf. 3.4).

Il existe cependant des études, mais en nombre limité, qui **analysent minutieusement les pratiques d'évaluation des acquis des étudiants**. Citons-en quatre, parmi celles qui seront le plus abondamment utilisées dans ce rapport. La première est une recherche de Warren Piper (1994), qui porte principalement sur les évaluateurs externes des universités anglaises (« external examiners »). Ce système original de contrôle externe de la qualité des pratiques d'évaluation est décrit en annexe (7.3). L'objectif principal de l'enquête de Warren Piper est d'analyser les rôles et les fonctions de ces évaluateurs externes, mais une première partie de la recherche vise à décrire finement les pratiques internes d'évaluation.

La deuxième étude a été produite par l'Agence de modernisation des universités⁸ qui a consacré une de ses Rencontres à « L'organisation et la préparation des examens ». Le compte rendu de cette journée (Poirier, 2001) fournit d'intéressants renseignements sur les pratiques d'évaluation des acquis des étudiants. Il montre en particulier combien la lourdeur de l'organisation des examens, accentuée par la semestrialisation des enseignements, a eu pour conséquence fâcheuse de reléguer au second plan les réflexions pédagogiques sur l'évaluation.

La troisième étude est une enquête fouillée sur les pratiques d'évaluation des apprentissages menée à l'Université de Montréal et dans ses écoles affiliées, par questionnaire auprès de 643 professeurs et par entretiens auprès de plusieurs échantillons d'enseignants et d'étudiants. Le rapport de recherche, publié en 1997, livre des indications précieuses tant sur les pratiques d'évaluation que sur les perceptions qu'en ont les enseignants et les étudiants (Blais et al., 1997).

Enfin, un petit nombre d'études récentes commencent à investiguer, dans le détail et de manière presque clinique, les comportements évaluatifs des enseignants du supérieur (York et al., 2000). Par la technique de la réflexion parlée ou par celle des entretiens de groupe, on cherche à savoir quels sont les comportements courants des enseignants du supérieur quand ils sont confrontés à des tâches authentiques d'évaluation. Par exemple, comment élaborent-ils leur jugement face à une copie ? Selon une démarche « platonicienne » de comparaison entre une copie modèle qu'ils ont en tête et les copies réelles ? Selon une approche intuitive, en regard de leur expérience antérieure de correcteur d'épreuves équivalentes ? En référence à une grille explicite de critères dont la pondération a été élaborée à l'avance ? S'appuient-ils sur un corrigé et si oui, quelle en est la nature ? Comment réagissent-ils face à un étudiant qui introduit dans sa réponse des éléments pertinents, mais qui n'étaient pas prévus dans leur corrigé ? ...

Quelles conclusions d'ensemble peut-on tirer de ces diverses études ?

3.1 UN ALLOURDISSEMENT DES TÂCHES D'ÉVALUATION, AYANT DES EFFETS NÉGATIFS SUR SA QUALITÉ

La première conclusion qui se dégage de ces études, confirmée d'ailleurs par les entretiens, est que l'évaluation des acquis des étudiants est devenue **une composante de plus en plus lourde du métier d'enseignant-chercheur**. Les tâches d'évaluation occupent une place grandissante au sein de leurs activités. Les enseignants déplorent que le temps consacré à l'évaluation grignote progressivement celui qui est dévolu à l'enseignement. La répartition entre 24 à 25 semaines d'enseignement et 12 à 13 semaines d'évaluation a souvent été évoquée. On la juge inadéquate, sans compter que des épreuves de contrôle continu sont aussi organisées au cours de la période dite d'enseignement. La massification, la modularisation et la semestrialisation sont souvent montrées du doigt comme étant les principales responsables de cette augmentation considérable du poids des examens.

Outre l'alourdissement des tâches d'évaluation qu'elle entraîne pour les enseignants-chercheurs⁹, cette évolution semble aussi avoir produit des **effets négatifs sur la qualité de l'évaluation** : régression, voire disparition des oraux ; diminution des possibilités laissées aux étudiants de choisir leur formule d'examens ; non-tenues de certaines réunions de jury, ... Poirier (2001) et Girod de l'Ain (1997) soutiennent que la massification a entraîné, en France, une diminution du contrôle continu et une modification de la nature même des épreuves, en particulier une régression des oraux, ce qui ne manque pas de modifier la nature des compétences développées (cf. 3.5). La multiplication d'examens déconnectés les uns des autres aboutit en effet au développement d'un « *travail intellectuel morcelé et concentré sur les prochains « obstacles » au détriment de la réflexion et de l'approfondissement, attributs fondamentaux d'études universitaires* » (Girod de l'Ain, 1997, p.108). Le rapport général du

⁸ Rebaptisée, depuis lors, en Agence de mutualisation des universités.

⁹ Alors que ceux-ci éprouvent déjà beaucoup de difficultés à gérer l'importante diversité des tâches concurrentielles qui sont liées à leurs différents « métiers » de chercheur, d'enseignant et de prestataire de services internes et externes.

CNE de 1995, dans sa partie consacrée aux difficultés du premier cycle massifié, signale aussi qu'un des effets les plus dommageables de l'augmentation des effectifs réside dans la modification de la pédagogie qu'elle implique nécessairement, en particulier en rendant impossible la correction fine d'une grande masse de copies et le maintien des examens oraux.

Blais et ses collaborateurs (1997) observent la même tendance au Québec : au premier cycle, plus les effectifs sont nombreux, plus l'examen final écrit constitue le mode d'évaluation privilégié¹⁰. Les modalités d'évaluation sont aussi davantage diversifiées pour un petit groupe (exposé, examen intermédiaire, travail de recherche, ...) que pour un grand groupe. Un lien entre la taille des groupes et le type d'habiletés mesurées semble, de plus, établi par leur enquête : les enseignants évaluent davantage la mémorisation au sein des grands groupes de premier cycle et cette habileté occupe ensuite une part de plus en plus ténue dans les examens des cycles supérieurs. Les enseignants soulignent aussi que l'accroissement du nombre d'étudiants rend plus problématique la réalisation d'une évaluation personnalisée.

Montgomery (1995) déplore, de la même manière, que la soudaine croissance des effectifs étudiants dans les universités anglaises ait entraîné le recul des méthodes d'évaluation traditionnelles qui nécessitaient un travail important de lecture et d'avis de la part des enseignants (essais, réponses longues personnelles, analyses de cas réels, ...) au profit de méthodes moins coûteuses, comme les QCM de connaissances corrigés mécaniquement¹¹.

En Allemagne aussi, les recommandations issues des recherches des années 70 concernant les examens universitaires n'ont pas été toutes mises en œuvre dans le courant des années 80, à la suite de la rapide expansion des effectifs. Ainsi, malgré le poids jugé excessif de l'oral terminal, le contrôle continu s'est peu développé, compte tenu de l'accroissement de la charge de travail qu'il aurait inévitablement entraîné dans l'université de masse (Kehm, 2001).

Bref, si les pratiques effectives d'évaluation sont encore assez mal connues, on sait au moins qu'elles représentent une part importante du travail professoral. On sait aussi que cette part aurait tendance à s'accroître, au grand dam des enseignants-chercheurs, suite à l'explosion des effectifs et aux réformes des structures de l'enseignement universitaire (semestrialisation et modularisation). On redoute enfin que cette évolution ait eu des conséquences importantes sur la qualité des pratiques d'évaluation, sur la nature des tâches proposées aux étudiants et, en définitive, sur le type de connaissances et de compétences qu'ils développent.

3.2 DES PRATIQUES D'ÉVALUATION PEU STANDARDISÉES

La deuxième conclusion majeure qui se dégage des études portant sur les pratiques d'évaluation dans le supérieur est la **grande hétérogénéité** de ces pratiques. Comme le souligne Pair (2001) dans son panorama des forces et des faiblesses de l'évaluation en France, la situation dans le supérieur contraste fortement avec les pratiques qui ont cours au premier et au second degré. À cause de la tradition d'autonomie des établissements d'enseignement universitaire (et, nous ajouterions, de la tradition d'autonomie des personnes au sein de ces établissements (cf. 3.2.2)), l'évaluation des acquis,

¹⁰ Les enseignants veillent cependant à ce que l'examen écrit final comporte des tâches complexes, comme la résolution de problèmes ou le développement long ou court. Ils n'attribuent généralement pas un poids démesuré aux tâches les plus simples (QCM, Vrai-faux). On note quand même que 9% de ceux qui ont recours aux QCM leur confèrent un poids supérieur à 80%.

¹¹ Sur le plus long terme, une étude passionnante a été consacrée aux facteurs qui expliquent pourquoi les examens oraux ont progressivement cédé la place aux écrits dans les universités d'Oxford et de Cambridge, durant le 18^e et le 19^e siècles (Stray, 2001). Cette étude montre notamment comment l'augmentation du nombre d'étudiants (déjà !), l'introduction de matières scientifiques et le type de critères d'évaluation (des critères socio-moraux aux critères de maîtrise d'acquis cognitifs individuels) ont contribué à cette lente modification de la nature des épreuves.

en ce compris la délivrance des diplômes, est de la compétence de chaque établissement et de chacun des enseignants, ces derniers évaluant leurs propres étudiants.

Bien sûr, des textes légaux encadrent l'évaluation des acquis dans l'enseignement universitaire et assurent un minimum de standardisation des procédures générales. Ainsi, une circulaire du 1^e mars 2000 relative à l'organisation des examens dans les établissements publics de l'enseignement supérieur français rappelle les réglementations en vigueur en matière d'examens universitaires. Concernant l'organisation des examens, la circulaire précise les règles qui régissent les aspects suivants : la convocation des étudiants, la constitution du jury, l'élaboration d'un procès-verbal pour chaque épreuve, les comportements à adopter en cas de fraude, la communication des résultats aux étudiants, les conditions de réussite de l'année et de l'obtention du DEUG et l'accès aux copies. Mais le texte de la circulaire insiste surtout sur la nécessité que chaque établissement définisse ses règles propres, les explicite et les fasse connaître dans la plus grande transparence et tôt dans l'année aux étudiants, étant donné qu'un grand nombre de décisions cruciales sont en fait prises au niveau de chaque établissement, voire de chaque unité d'enseignement. Ainsi, la loi du 26 janvier 1984 ne trace qu'un cadre très général et pour le moins ouvert sur le rapport entre évaluation continue et terminale : *« les aptitudes et l'acquisition des connaissances sont appréciées, soit par un contrôle continu et régulier, soit par un examen terminal, soit par ces deux modes de contrôle combinés »*. Reste que la formule permet localement de nombreux cas de figure... De plus, ces textes définissent les règles générales à suivre quant à l'organisation des épreuves, mais ils ne disent rien de leur contenu.

Les pratiques d'évaluation se différencient d'abord en regard de **l'importance relative** qu'elles accordent au **contrôle continu** par rapport au contrôle terminal et en regard de la **nature** des épreuves de contrôle continu. Selon l'étude de Warren Piper (1994), l'évaluation des étudiants du supérieur peut se réaliser sur la base de trois sources possibles, combinées, à des degrés très divers, selon chaque enseignant :

- des performances que l'étudiant produit dans des conditions standardisées d'examen (examens écrits, oraux, par QCM, en laboratoires, ...)
- des performances que l'étudiant produit en dehors de conditions standardisées d'examen (essai, dissertation, travaux, rapports, compte rendu d'expériences, ...)
- des rapports que des personnes extérieures ont rédigés à propos des performances de l'étudiant (maître de stage, tuteur, ...).

Selon la même étude, une partie importante des enseignants anglais (82%) établissent leur évaluation en y intégrant désormais un contrôle continu, réalisé en dehors ou avant l'examen final, contrôle continu dont les examinateurs externes éprouvent davantage de difficultés à contrôler la validité. Ce travail de l'étudiant extérieur à l'examen final proprement dit intervient, en moyenne, pour près de 28% de la note (14% étant évalués avant l'examen et les autres 14% étant du travail effectué en dehors de l'examen mais présenté lors de celui-ci). Au Royaume-Uni, on estime souvent que l'explication la plus plausible de l'augmentation continue des mentions et de l'inflation des grades réside dans l'importance qu'ont progressivement pris les travaux effectués par l'étudiant à domicile dans le jugement de l'enseignant, ces derniers étant généralement notés plus favorablement que les copies d'examen réalisées sans préparation (Elton, 1998 ; Gibbs & Lucas, 1997). Ce changement dans les pratiques d'évaluation repose la question des standards nationaux et de leur signification à travers le temps (cf. 4.1).

Comme nous l'avons noté au point 3.1, la part relative du contrôle continu dans l'évaluation semble tributaire de la taille du groupe d'étudiants : plus les étudiants sont nombreux, plus les enseignants réduisent la part du contrôle continu, face à la lourdeur des corrections (Gibbs & Lucas, 1997). Or, les travaux exigés des étudiants tout au long du cours semblent garantir un certain niveau dans la qualité des apprentissages : ils incitent les étudiants à privilégier le travail régulier plutôt que le bourrage de crâne de dernière minute, ils leur fournissent des feedbacks formateurs, ... Les résultats des étudiants au contrôle continu semblent d'ailleurs, davantage que leurs résultats aux examens,

corrélés aux mesures d'acquis effectuées à long terme, sans doute parce qu'ils se rapprochent des tâches qu'ils seront amenés à exécuter dans leur vie professionnelle.

Une autre dimension de l'hétérogénéité des pratiques d'évaluation a trait aux **spécificités disciplinaires** : l'évaluation des acquis se réalise différemment selon la filière d'études (Brown & Glasner, 1999 ; Heywood, 2000 ; Warren Pippet, 1994). Ainsi, les épreuves de calcul numérique semblent plus fréquentes en sciences, les travaux pratiques aussi. Le recours à des performances non langagières s'observe surtout en faculté d'ingénieur et de technologie. Plus curieusement, l'examen oral semble privilégié en sciences. Les enseignants de ces filières estiment qu'il s'agit d'une méthode valide qui permet de discriminer fortement les étudiants, sans doute parce que l'oral permet de reposer des questions précises et de « tester » ainsi la compréhension fine de concepts spécifiques. À l'inverse, les enseignants des facultés des sciences humaines et sociales et de lettres émettent des doutes sur la validité de cette forme d'examen et privilégient des épreuves écrites à réponses ouvertes (l'essai, la dissertation et les questions à développement long ou court).

Les spécificités disciplinaires s'observent jusque dans les distributions des notes des étudiants : ces distributions sont en effet différentes d'une discipline à l'autre, en termes de moyenne, d'écart-type et d'allure générale de la courbe (Bridges et al, 1999 ; Yorke et al., 2000). Ainsi, les disciplines scientifiques ont tendance à recourir à une large palette de notes alors que l'empan des notes est considérablement plus restreint dans les autres filières. En conséquence, la proportion des mentions y est dès lors aussi fort différente : ainsi, dans les universités anglaises, alors que 21% des étudiants en mathématiques obtiennent la meilleure mention, ils ne sont que 3,7% à y avoir accès dans les études juridiques. Cette différence soulève un problème d'équité entre étudiants de différentes filières : ils n'ont pas, au départ, la même probabilité d'obtenir des mentions. En cherchant à comprendre ce qui pourrait expliquer ces conventions implicites de notation spécifiques à chaque filière, Bridges (1999) montre que la nature des connaissances et des compétences qui font l'objet de l'évaluation peut être mise en parallèle avec l'usage de l'échelle de notation. En histoire par exemple, les enseignants déclarent chercher à mesurer la maîtrise de démarches assez générales, telles que la critique des sources. Il leur semble que les extrémités de l'échelle constituent des « territoires peu sûrs », qu'ils préfèrent dès lors ne pas explorer, l'excellence et, à l'autre bout de l'échelle, la médiocrité dans la maîtrise de ces démarches étant rarement atteintes. Par contre, des enseignants de mathématiques et d'informatique déclarent procéder à une addition de réponses correctes ou incorrectes et s'autorisent alors à recourir à l'ensemble des échelons de l'échelle de notation, y compris dans ses deux extrémités.

L'absence de standardisation des pratiques d'évaluation s'observe aussi à **l'intérieur d'une même filière**, même dans le cas de diplômes nationaux. L'étude de cas, rapportée en annexe (7.1), a cherché à explorer les spécificités des pratiques d'évaluation de deux universités, pour une même filière. Les rapports transversaux du Conseil national d'évaluation fournissent de belles illustrations de l'hétérogénéité des pratiques¹². À titre d'exemple, le rapport portant sur les formations offertes par les 24 UFR de pharmacie aboutit à la conclusion suivante, en ce qui concerne le contrôle des connaissances : « *chaque UFR organise librement, dans le cadre de son autonomie, la préparation au concours et décide des épreuves, ce qui pose la question de l'homogénéité des résultats* » (CNE, 1998a, p. 34).

L'étude de Jarousse et Michaux (2001) montre également que les modalités d'évaluation varient considérablement d'une filière à l'autre, mais aussi d'une université à l'autre pour une même filière. D'après leur enquête auprès de 155 responsables de DEUG, les pratiques d'évaluation des acquis se différencient sur les points suivants.

¹² Même si les pratiques d'évaluation des acquis n'occupent pas une place très importante dans les critères retenus par cet organisme pour évaluer les universités et les formations. À titre d'exemple, le volume 2 du rapport concernant les formations de pharmacie, qui dissèque les points forts et les points faibles des 24 UFR, ne consacre que quelques lignes, parfois deux seulement, à cette question, sous la rubrique « les examens et les oraux ». De plus, il arrive qu'au sein de ces quelques lignes soit aussi abordée la question de l'évaluation de l'enseignement (CNE, 1998b).

- Les conditions générales de certification¹³ : certification annuelle avec capitalisation modulaire (avec ou sans condition de moyenne et compensation entre les modules), capitalisation modulaire avec compensation à l'issue des deux années, validation semestrielle des acquis, ...
- Les conditions de rattrapage : place de la session de rattrapage (en septembre ou juste après la session terminale), modalités de conservation des notes (condition de moyenne, choix de la note conservée (la meilleure des deux notes, celle de la session de rattrapage), ...).
- La pondération des modules et du contrôle continu. Ici aussi, c'est la diversité qui domine. Le contrôle continu se voit accorder une importance tantôt proportionnelle à sa part dans le volume d'enseignement, tantôt inférieure, tantôt supérieure. Son poids dans l'évaluation peut ainsi passer de 35% dans une filière à 59% dans une autre. Dans le même sens, le poids des modules obligatoires et optionnels est aussi très variable : pour une même filière, ils ont un poids équivalent dans tel établissement, alors que les modules obligatoires ont un coefficient de 3 dans tel autre.

En bout de course, l'évaluation des acquis des étudiants apparaît donc comme entièrement décentralisée : « aucune épreuve commune ne permet de comparer sur une base externe ce que les étudiants ont effectivement appris durant leur cursus » (Duru-Bellat et al., 2000, p. 138). Il en découle que la définition même de la réussite est contextualisée : « Tant les notes que les décisions de passage (...) sont entièrement fabriquées au sein de chaque filière » (idem, p.138). Chaque équipe d'enseignants développe des pratiques d'évaluation spécifiques, qui tiennent probablement compte des caractéristiques des étudiants accueillis et qui visent implicitement à respecter certaines normes docimologiques tacites, par exemple une moyenne du groupe d'étudiants qui ne soit ni trop basse ni trop élevée (cf. point 3.3).

Bien sûr, il n'y a pas de règle sans exception. L'Association nationale des directeurs de maîtrise en sciences de gestion a institué, depuis plusieurs années, des épreuves communes d'admission à la Maîtrise en Sciences de gestion. Ces épreuves, dénommées Message¹⁴, constituent le socle commun des examens d'admission pratiqués au sein des différents établissements, chacun de ces établissements disposant de la liberté d'ajouter localement des épreuves spécifiques. Un des objectifs affichés de cette collaboration est d'assurer une plus grande homogénéité des procédures de recrutement.

Plusieurs documents consultés s'interrogent sur les **multiples causes de cette absence de standardisation des pratiques d'évaluation** des acquis. Citons-en quelques-unes, parmi les principales.

3.2.1 Absence de standardisation des objectifs, des contenus et des méthodes

Le premier facteur explicatif du manque de standardisation des pratiques d'évaluation réside bien sûr dans l'absence de standardisation des objectifs, des contenus et des méthodes des formations elles-mêmes. Autrement dit, si les acquis des étudiants sont évalués de manière très diversifiée, c'est parce que les compétences attendues aux différents paliers de la formation n'ont pas fait l'objet d'une définition explicite, précise et standardisée. En quelque sorte, l'absence de données à propos des acquis renvoie ici à la question de la grande diversité des formations tertiaires elles-mêmes. Dit plus crûment, si nous ne savons pas ce que les étudiants ont appris, c'est notamment parce qu'ils ont appris, pour un même intitulé de diplôme, des choses extrêmement différentes selon l'établissement dans lequel ils ont poursuivi leurs études.

Ce phénomène s'explique par la très large autonomie dont jouissent les établissements et les enseignants du supérieur quant à la définition du contenu précis des cursus. En France, l'autonomie croissante dont ont bénéficié les universités en matière d'organisation pédagogique et les effets de

¹³ Une importante hétérogénéité préside aussi, au niveau international, au mode de certification dans le supérieur, tout au long d'un continuum allant de la conception additionniste des crédits des études graduées américaines jusqu'au modèle allemand des épreuves finales uniques (Girod de l'Ain, 1997).

¹⁴ Ces épreuves sont décrites sur le site <http://www.msgfrance.org/message.htm>.

cette autonomie ont été particulièrement bien décrits récemment par Musselin (2001), Jarousse et Michaut (2001) et Duru-Bellat, Jarousse, Leroy-Audouin et Michaut (2000).

Même si une composante importante de l'enseignement supérieur français réside dans la volonté de « *construire un système national qui, dans une version idéale, reproduirait à l'identique les mêmes institutions, les mêmes cursus et les mêmes enseignants sur l'ensemble du territoire* » (Musselin, 2001, p. 79), l'évolution de la population étudiante et l'émiettement des formations ont abouti à douter sérieusement de l'équivalence des titres nationaux. Dans sa description de la lente émergence des universités comme entités intermédiaires entre l'État et les facultés, Musselin (2001) analyse comment, malgré les deux principes d'uniformisation et d'égalitarisme reconnus par la loi¹⁵ et auxquels adhèrent les acteurs de l'université, les formations sont devenues, dans les faits, de plus en plus hétérogènes. La palette des formations s'est considérablement étendue, le public lui-même est devenu plus hétérogène et surtout, pour s'y adapter, les pratiques locales se sont diversifiées. Musselin (2001) décrit aussi comment les autorités de tutelle ont cherché en vain à concilier l'augmentation sans précédent du nombre de cursus et la garantie d'équivalence entre ceux-ci. Ni la multiplication des maquettes en proportion des nouveaux cursus, ni la définition très ouverte de ces maquettes¹⁶ ne constituent des solutions satisfaisantes. Dans le dernier cas en particulier, la maquette est définie de manière tellement large qu'elle n'offre plus la garantie d'équivalence pour laquelle elle avait été créée : « *dans un cas comme dans l'autre, le maintien des maquettes nationales est peu compatible avec la diversité croissante des filières et rend de plus en plus illusoire la garantie de l'homogénéité des cursus d'un bout du territoire à l'autre* » (Musselin, 2001, p.100).

Jarousse et Michaut (2001) ont aussi mis en évidence l'hétérogénéité considérable des modes d'organisation pédagogique des premiers cycles, qui se différencient notamment sur les trois dimensions suivantes.

- L'organisation générale du premier cycle : présence ou non d'une phase d'accueil, contenu de cette phase, volume annuel d'enseignement, proportion des divers types d'enseignements (magistral, TP, TD, ...), statut des enseignants.
- Le curriculum : proportion d'enseignements optionnels (de 0 à 10% pour une même filière) ; spécialisation et pré-professionalisation précoces ou retardées ; importance accordée aux différentes disciplines, notamment en termes de volume horaire.
- Les dispositifs d'aide à la réussite : forme des tutorats et volume horaire associé, forme des remises à niveau et volume horaire associé, fréquence des permanences, ...

Une conséquence importante de cette grande diversité réside dans le fait que les étudiants inscrits dans une même filière sont confrontés à des conditions d'enseignement très différentes selon l'université fréquentée. Il est dès lors probable qu'ils y développent des compétences et des connaissances différentes.

Dans ce domaine aussi, les rapports transversaux du Conseil national d'évaluation sont très instructifs. Par exemple, celui qui porte sur les formations de pharmaciens relève que, même pour ce diplôme national avec concours, l'arrêté qui fixe le contenu des cursus, s'il évoque bien « *une mosaïque de plus d'un millier de notions qui doivent être obligatoirement étudiées* », ne précise « *ni le poids relatif, ni le poids absolu des différentes matières ou disciplines, ni la chronologie des*

¹⁵ Comme le rappelle le récent arrêté du 8 avril 2002 relatif aux grades, aux titres universitaires et aux diplômes nationaux, ces derniers « *confèrent les mêmes droits à tous les titulaires, quels que soient les établissements qui les ont délivrés et les modes d'acquisition* », même si ce sont les différents établissements qui, jouissant de la liberté pédagogique et scientifique, « *sont autorisés à délivrer, au nom de l'État, les diplômes nationaux* ».

¹⁶ Elles se contentent en effet, pour les diplômes nationaux, de préciser le nombre minimal d'heures de cours, les règles générales de l'évaluation des connaissances et une liste de sujets obligatoires à traiter (Duru-Bellat, Jarousse, Leroy-Audouin et Michaut, 2000).

enseignements » (CNE, 1998a, p.36). Le rapport regrette aussi que les objectifs soient assez mal définis et le soient en termes trop généraux. Une conséquence importante est que les solutions retenues par les différentes UFR pour poursuivre une même finalité peuvent varier considérablement, au point sans doute de ne pas assurer les mêmes acquis à tous les étudiants. Ainsi, pour développer les capacités de synthèse de l'étudiant, telle UFR multiplie les renvois bibliographiques en maintenant le volume horaire constant, telle autre procède à une diminution de la charge horaire de 20% et introduit des enseignements thématiques (CNE, 1998a, p.37).

Le rapport Fitoussi (2001) sur l'enseignement supérieur de l'économie évoque, lui aussi, cette question de l'absence de standardisation des niveaux d'études. Le rapport signale que la variété des spécialisations de deuxième et de troisième cycle n'est pas en soi condamnable, si l'on veille en parallèle à conserver la cohérence de la formation. Bien au contraire, cette diversité reflète le souci des universités de tirer le meilleur parti de leurs ressources humaines. Pour le premier cycle par contre, alors qu'aucune mention de spécialité ne figure sur le diplôme et qu'un cadre est fixé par le Ministère, le rapport s'étonne de la grande hétérogénéité des cursus. Ainsi, le volume horaire de l'enseignement des mathématiques varie, sur les deux années du DEUG, de 95,5 à 245 heures, soit un rapport de 1 à 2,5. Certaines formations sont absentes de certains cursus de premier cycle, et non des moindres en regard d'une formation à l'économie, comme l'informatique ou la comptabilité. Le rapport n'évoque pas ce qui est effectivement enseigné sous ces intitulés. Sans s'avancer beaucoup, on peut supposer que la diversité des contenus et des méthodes accroît encore l'hétérogénéité des cursus et donc celle des pratiques d'évaluation. Au rang des dysfonctionnements, le rapport signale d'ailleurs que les finalités des différents cursus n'apparaissent pas toujours clairement et que les savoirs qui devraient être acquis à telle ou telle étape de ces cursus ne sont pas formellement identifiés. Dans ses propositions, le rapport suggère que les exigences du premier cycle soient mieux explicitées et fournit, à titre d'exemple, une définition par compétences des objectifs de formation d'un premier cycle anglais d'économie.

Le rapport Dejean (2002), consacré à l'évaluation de l'enseignement, met lui aussi en évidence l'absence de standardisation des contenus et des objectifs des diplômes, même dans le cas où ceux-ci reçoivent une habilitation nationale. La notion même de diplôme national serait une pure fiction, tant les procédures par lesquelles l'État apporte une garantie d'équivalence aux formations locales sont lâches et sommaires.

Cette tendance à douter de l'équivalence des titres nationaux s'observe aussi au Royaume-Uni. Selon le rapport de Warren Piper (1994), la large autonomie accordée aux universités dans l'évaluation des acquis des étudiants remet en cause l'équivalence de niveau des différents diplômes (cf. 4.1.1). Celle-ci est décrite comme une « fiction polie », même si des instituts de contrôle de la qualité sont censés la garantir. Une tension est vécue entre des exigences de diversité des programmes (notamment pour répondre à des publics variés) et des exigences d'équité.

3.2.2 Une longue tradition de « liberté académique »

Un deuxième facteur explicatif de l'absence de standardisation des pratiques d'évaluation des acquis réside dans la grande autonomie dont jouit, en sus de l'autonomie institutionnelle, chaque enseignant-chercheur dans l'organisation de son enseignement et dans la mise au point du contenu de ses examens. Quand bien même la description officielle d'un même curriculum serait semblable dans deux établissements, les enseignants à qui sont attribués les cours se retrouvent pratiquement libres de déterminer leur contenu et leur forme. Le curriculum réel peut dès lors être très différent, pour un même descriptif officiel de cours, comme le montre l'étude de cas (7.1). La « liberté académique » n'est pas un vain mot. Dans la conception humboldtienne de l'université moderne, la notion de liberté académique représente une exigence d'autonomie institutionnelle de l'Université, d'indépendance de cette dernière par rapport à l'État et aux autorités publiques. Cette autonomie est considérée comme une condition indispensable au développement d'une recherche de qualité et créative, parce que cette dernière peut alors se déployer hors des contraintes limitantes et à court terme du pouvoir (Renaut, 1995). Cette exigence d'autonomie s'étend aussi à l'individu, qui doit pouvoir, au sein de l'Université, poursuivre sa recherche « *dans la solitude et la liberté* », « *sans qu'une contrainte ou un but déterminé* »

lui soit imposé » (Renaut, 1995, p. 130). On comprend bien comment cette tradition de liberté académique a laissé de profondes traces, y compris en qui concerne la liberté individuelle revendiquée par chaque enseignant d'organiser son enseignement et ses examens comme il l'entend. Pour reprendre une image évoquée lors d'un colloque sur l'enseignement supérieur, l'université est un temple moderne du savoir libre et gratuit, les professeurs en sont les grands prêtres et personne ne doit s'occuper de la manière dont ils disent la messe... Pour caricaturale qu'elle soit, cette métaphore constitue une bonne image du culte de la liberté académique, au nom de laquelle chaque enseignant serait une personnalité libre, unique, non interchangeable, qui ne doit subir aucune contrainte dans sa création de savoir mais aussi dans la manière d'organiser son enseignement (Dejean, 2002 ; Lahire, 1997).

3.2.3 Une grande diversité de conceptions de l'évaluation

Un troisième facteur explicatif de la grande diversité des pratiques d'évaluation est que l'enseignant-chercheur, puisqu'il dispose d'une importante marge d'autonomie, organise l'évaluation des acquis de ses étudiants selon son profil personnel d'évaluateur. Des recherches récentes ont en effet mis en évidence que les **conceptions qu'ont les enseignants universitaires de l'évaluation** pouvaient être très différentes. Samuelowicz & Bain (2002), par exemple, ont établi trois profils d'évaluateur.

- Dans le premier de ces profils, le professeur cherche surtout à **mesurer la capacité de l'étudiant à reproduire de l'information**, telle qu'elle a été présentée au cours et/ou dans le photocopié. L'évaluation est considérée comme un moyen de pression pour faire travailler les étudiants. La différence entre les étudiants qui réussissent et ceux qui échouent tient dans la quantité de matières qu'ils sont capables de restituer correctement.
- Le second profil rassemble les examinateurs qui cherchent à mesurer la **capacité de l'étudiant à reproduire de l'information et à l'appliquer à des situations nouvelles**. La compréhension implique donc, de leur point de vue, une transformation ou une réinterprétation des connaissances acquises de manière à répondre aux exigences d'une tâche nouvelle.
- Au sein du troisième profil, les examinateurs conçoivent des épreuves qui mesurent la **capacité de l'étudiant à intégrer, transformer et utiliser de manière personnelle des connaissances**.

3.2.4 Une mixité de fonctions difficilement conciliables

L'évaluation peut avoir plusieurs fonctions. À côté des deux fonctions traditionnelles de sélection et de certification qui ont prévalu tant que l'université était une formation réservée à une élite et que son rôle social était donc de sélectionner et de « valider » cette élite, la fonction formative de l'évaluation, c'est-à-dire d'aide à l'apprentissage, a été progressivement mise en avant dans l'université de masse. Comme le signale un document de politique académique de l'Université de Sherbrooke au Canada, *« l'évaluation dépasse le simple geste « sanctionnel » : elle est une composante indissociable de la dynamique même de l'apprentissage »*. Cette nécessité de ne pas dissocier l'examen de la formation incite notamment à accorder davantage d'importance au contrôle continu, afin que l'évaluation joue un rôle central dans la progression de l'étudiant (Poirier, 2001 ; Greer, 2001 ; Knight, 2002 ; Langevin & Bruneau, 2000)¹⁷.

La mise en œuvre simultanée de diverses fonctions de l'évaluation au sein des épreuves de l'enseignement supérieur est sans doute de nature à accroître encore l'hétérogénéité des pratiques d'évaluation. De plus, certaines difficultés rencontrées actuellement par les systèmes traditionnels d'examens dans le supérieur s'expliquent par le fait qu'ils remplissent, dans les faits, plusieurs fonctions, alors que ces dernières sont parfois inconciliables au sein des mêmes épreuves. Prenons un

¹⁷ Même si les étudiants eux-mêmes, conditionnés par des pratiques d'évaluation centrées sur le seul résultat, se montrent peu sensibles à la dimension formatrice de l'examen : par exemple, 16% seulement d'entre eux considèrent l'évaluation comme une occasion de réflexion sur leurs apprentissages (Navarro, 2002).

exemple issu du compte rendu de la journée d'étude de l'AMU consacrée à « L'organisation et la préparation des examens » (Poirier, 2001). Ce compte rendu relève que les étudiants souhaitent obtenir leurs notes très rapidement après la première session, dans une perspective formative (s'améliorer pour la seconde, voire se réorienter), alors que la session d'examen, conçue dans la perspective traditionnelle de certification, s'efforce de répondre aux exigences docimologiques requises pour une telle évaluation (représentativité des questions, anonymat des corrections, ...). Or, ces exigences rendent difficile la communication rapide des résultats. Dans le même sens, une évaluation réalisée dans un but de sélection requiert que les épreuves soient fortement discriminantes, quitte à réduire considérablement le champ des compétences testées, alors qu'une évaluation qui cherche à dresser un bilan des acquis veille surtout à couvrir l'ensemble des compétences visées, quitte à ne pas discriminer finement les performances des individus.

Austin (1991) signale d'ailleurs que la fonction la plus traditionnelle de l'évaluation dans le supérieur reste sa capacité très classique à inciter l'étudiant à travailler. L'évaluation devient alors un système de récompense et de punition, de « carotte et de bâton », susceptible d'instrumentaliser le travail étudiant. Aux yeux des étudiants, l'évaluation possède un pouvoir incitatif indéniable : elle constitue une puissante motivation externe à leur travail. Chaque enseignant réclame « son » examen : pour assurer son pouvoir sur les étudiants, il a avant tout besoin de la possibilité de les faire échouer¹⁸. L'examen est, dans le même temps, exigé par les étudiants, comme récompense du travail fourni, il est revendiqué comme un droit inaliénable. Ainsi, les étudiants sont très attachés au principe d'organiser des examens sur tous les éléments du programme. Ils s'opposent de manière catégorique au tirage au sort des examens, car ils se sentiraient frustrés de ne pas présenter une épreuve sur une matière dans laquelle ils se sont investis (Poirier, 2001). Tout travail mérite salaire : tout enseignement doit donc donner lieu à une note. Selon Austin, une évolution majeure des pratiques d'évaluation consistera à passer de cette conception classique à une conception davantage centrée sur l'évaluation comme moyen de retourner de l'information à l'étudiant pour assurer ses apprentissages. Il cite, à titre d'exemple, l'expérience de l'Université de Californie à Santa Cruz qui ne délivre plus de notes, mais uniquement des commentaires qualitatifs écrits qui renseignent les étudiants sur la qualité de leur travail et sur la manière de l'améliorer.

3.3 UNE ÉVALUATION NORMATIVE, QUI NE SE FONDE PAS SUR DES OBJECTIFS EXPLICITES DE FORMATION

L'hétérogénéité des pratiques d'évaluation n'est peut-être pas condamnable en soi, si l'on met de côté les problèmes d'équité et d'équivalence des titres. Elle pourrait même se justifier, nous l'avons vu, en termes d'adaptation pédagogique aux caractéristiques des « nouveaux » étudiants que l'université de masse a en charge d'accueillir ou en termes d'adéquation des compétences de sortie au contexte socioprofessionnel régional. Si le fait d'évaluer, pour une même filière, des acquis différents au sein d'établissements autonomes pourrait donc être encore acceptable, on voit mal ce qui pourrait justifier le fait de ne pas savoir précisément quels acquis font l'objet de l'évaluation. L'absence d'explicitation de ce que l'on mesure est en effet gênante à plus d'un titre : les étudiants manquent de repères sur les manières d'étudier les plus propices à les préparer aux examens et les futurs employeurs ne disposent pas d'inventaires précis des compétences de ceux qui se présentent sur le marché de l'emploi. Ces nombreux effets dommageables seront décrits au point 4.2.

Par ailleurs, l'explicitation des objectifs de formation est, en principe, au cœur du processus évaluatif : évaluer revient à recueillir de l'information de manière systématique, fidèle et valide pour prendre des décisions, en particulier pour juger de l'atteinte des objectifs de formation par les étudiants. L'évaluation suppose donc que les objectifs des cursus aient été préalablement explicités en termes de connaissances et de compétences, pour qu'elle puisse se réaliser en regard de ces objectifs. Or, la troisième conclusion majeure qui s'impose à la lecture des études sur les pratiques d'évaluation

¹⁸ Cet « attachement viscéral » de chaque enseignant-chercheur à disposer de l'arme de l'examen constitue un des freins majeurs à l'organisation d'épreuves intégrées, communes à un groupe d'enseignements.

est que le **système des examens universitaires fonctionne la plupart du temps en l'absence de définition, même locale, des connaissances et des compétences qui sont attendues en fin de formation.**

Le type d'évaluation qui se développe dans l'enseignement universitaire est essentiellement normatif : l'évaluation s'attache principalement à classer les étudiants les uns par rapport aux autres. L'attribution contextualisée des notes et le caractère local des décisions de réussite participent au développement d'une évaluation qui ne cherche pas prioritairement à mesurer les compétences acquises par les étudiants en fonction d'objectifs explicites de formation, mais bien à les distinguer les uns des autres¹⁹. L'ajustement local des épreuves permet ainsi aux enseignants de conserver, d'année en année, une distribution de leurs notes jugée acceptable. Ce phénomène a bien été décrit au premier et au second degrés sous le nom d'effet « Posthumus » : un enseignant a tendance à ajuster le niveau de son enseignement et de son évaluation de façon à conserver d'année en année approximativement la même distribution gaussienne de ses notes (Crahay, 1996). Sans doute cet effet est-il également à l'œuvre dans le supérieur, avec des normes implicites différentes, cela va sans dire. Ainsi, une distribution jugée acceptable en première année du premier cycle se rapproche davantage de la courbe en i que de la distribution normale...

Lors des entretiens et de l'étude de cas, des pratiques d'évaluation ouvertement normatives ont été rapportées. Ainsi, il est fréquent que des notes provisoires soient d'abord attribuées à une série de copies, notes qui sont ensuite adaptées selon les résultats de l'ensemble du groupe, de manière à aboutir à cette fameuse norme implicite. Autre exemple, la correction s'effectue parfois en recherchant d'abord une très bonne et une très mauvaise copie puis en ajustant ensuite les autres sur ces deux extrêmes. Plus clairement encore, des programmes informatiques de corrections automatiques de questionnaires à choix multiples permettent d'ajuster le tarif de manière à obtenir la moyenne souhaitée. L'enquête de Blais (1997) rapporte aussi le souci de certains enseignants de rechercher une « moyenne historique, raisonnable ».

À côté des problèmes d'équité que pose l'évaluation normative (les standards de qualité selon lesquels sont jugés les étudiants dépendent du groupe dans lequel ils se trouvent), cette évaluation pêche aussi par un manque de transparence sur les acquis. En effet, elle permet juste de savoir quels étudiants ont appris le plus ou le mieux, mais l'évaluation normative ne nous renseigne pas sur ce qu'ils ont acquis, ni même sur le fait qu'ils aient appris beaucoup ou peu. D'ailleurs, les gains en termes de connaissances (par rapport à leur niveau de départ) des étudiants en échec se révèlent parfois identiques à ceux des étudiants qui obtiennent des mentions (Astin, 1991). L'évaluation normative fournit plutôt une indication sur la manière dont les étudiants se classent les uns par rapport aux autres, à un moment donné.

Un autre reproche que l'on peut adresser à l'évaluation normative est qu'elle crée ce que les économistes appellent un « bien rare », puisque seul un pourcentage minime et prédéfini d'étudiants se trouveront en tête du groupe, indépendamment des efforts produits par les étudiants et de l'importance de leurs acquis. La rareté exacerbant l'importance d'un bien, les performances moyennes sont alors considérées comme médiocres : l'évaluation normative garantit en quelque sorte qu'une majorité d'étudiants ne soient pas satisfaits de leurs résultats...²⁰

Comment expliquer que l'évaluation des acquis soit essentiellement normative et se réalise peu en regard d'objectifs de formation ?

¹⁹ La terminologie des mentions en témoigne : la « grande distinction » est attribuée aux quelques rares étudiants qui « sortent du lot ».

²⁰ À moins bien sûr de céder à l'inflation des grades et des mentions, comme on l'observe au deuxième et au troisième cycles pour cette même raison.

3.3.1 Des formations qui ne sont pas définies en termes de compétences à acquérir

Le premier facteur explicatif et le plus évident est que les formations elles-mêmes ne sont pas prioritairement conçues autour d'objectifs explicites d'acquisition de connaissances et de compétences. L'absence de données à propos des acquis, mise en évidence au point 2, renvoie ici à la question de savoir quels acquis devraient être évalués et donc quels sont les objectifs assignés aux formations. Dit plus abruptement, si nous ne savons pas ce que les étudiants ont appris dans leur cursus, c'est notamment parce que nous ne savons pas ce qu'ils auraient dû y apprendre...

Pour comprendre cet état de fait, il faudrait analyser plus longuement les modes de constitution des programmes de l'enseignement supérieur. De nombreux indices incitent à penser que la confection des programmes (le choix des matières, leur contenu, leur articulation et leur répartition horaire) est souvent plus « tributaire de la configuration des opportunités et des intérêts internes à l'institution » (Hutmacher, 2001, p.43) que d'un inventaire méthodique des compétences à développer, par exemple sur la base d'une analyse des conditions d'exercice des métiers auxquels donne accès le diplôme. Comme le signale cet auteur, il s'agit sans doute d'un des obstacles majeurs à une véritable professionnalisation des formations supérieures.

Cette tendance à construire des plans de formation selon les opportunités, voire selon les personnes (leur profil, leurs intérêts de recherche, leur champ de compétences, etc.) qui en auront la charge a été particulièrement bien décrite par Kletz et Pallez (2001) dans l'étude qu'ils ont consacrée aux mécanismes de constitution de l'offre de formation des établissements universitaires. Ils montrent bien comment la « spirale inflationniste » de l'offre de formation que l'on observe actuellement est la conséquence du mode de construction de l'offre de formation. La logique individuelle des enseignants-chercheurs les incite à multiplier des offres de formation en regard de leur champ de recherche et en fonction de stratégies personnelles (attacher son nom à un diplôme, assurer l'obligation de remplir sa charge d'enseignement, ...). Cette logique personnelle n'est contrecarrée ni par des stratégies d'établissement ni par des procédures d'expertise au niveau national. Pour notre propos, il faut retenir de cette étude que la création d'offres de formation trouve plus fréquemment son origine dans une initiative personnelle, liée parfois à des enjeux de carrière strictement individuels, que dans une analyse détaillée des besoins des formés ou encore un inventaire des compétences à développer chez ces formés en regard d'inadéquations qui auraient été observées entre leurs acquis et les besoins du marché du travail²¹.

D'ailleurs, la logique qui préside encore de manière dominante à la confection des programmes de formation universitaire est une logique de contenus, interne au monde académique. Un programme reste un assemblage de matières et de cours, une mosaïque dont le découpage et l'agencement sont liés aux disciplines scientifiques de référence et à leur mode de structuration. Ainsi, le programme d'histoire est articulé autour des grandes spécialités reconnues qui ordonnent ce champ scientifique et la communauté de ses chercheurs. Une étude portant sur les processus de développement de nouveaux cursus confirme la prédominance de cette logique de discipline : « *les nouveaux projets étaient toujours lancés à l'initiative d'un universitaire ou d'un petit noyau de collègues, en fonction de leurs propres préoccupations et intérêts* » (Mignot-Gérard et Musselin, 2002, p. 12). Concevoir des formations en termes d'acquis représenterait une tout autre entrée : qu'est-ce que les étudiants doivent savoir et savoir faire à la sortie du programme ? Quel type de connaissances et de compétences cherche-t-on à leur faire mobiliser et face à quel type de situations ?

Un des résultats des modes actuels de constitution des plans de formation réside dans l'aspect « patchwork » des programmes universitaires : l'éparpillement des enseignements permet

²¹ Même si cette même étude montre que le critère « débouchés » prend progressivement une place majeure dans l'évaluation des dossiers. Mais, même dans ce cas, il s'agit la plupart du temps d'un « faux-semblant », les justifications en termes de débouchés ne se basant que sur des descriptions peu formalisées des métiers susceptibles d'être concernés et sur des inventaires sommaires des besoins.

difficilement aux étudiants de les insérer dans une démarche globale (Girod de l'Ain, 1997 ; Lyon-Caen, 2002).

Le fait que les formations ne soient pas construites autour d'une explicitation de compétences à acquérir mais qu'elles s'élaborent selon des logiques internes à l'institution, selon son histoire²², voire selon ses stratégies de positionnement concurrentiel, apparaît aussi dans de nombreux passages des rapports du CNE. Ainsi, l'évaluation des différentes UFR de pharmacie (CNE, 1998b) signale l'existence d'un déséquilibre dans la formation dispensée par telle unité, déséquilibre attribuable au fait que les enseignements des disciplines dont cette unité est spécialiste sont privilégiés au détriment d'autres disciplines. Un rapport d'évaluation institutionnelle d'une université déplore l'absence de réflexion sur l'offre de formation et conclut que « *l'offre de formation apparaît ainsi davantage comme la résultante d'une course pour augmenter au maximum les effectifs étudiants, que comme l'aboutissement d'une politique mûrie et réfléchie* » (CNE, 2001, p.74). Tel autre rapport institutionnel souligne le caractère « anarchique » du processus de développement de l'offre de formation au sein d'une université, mais tempère son jugement en rappelant qu'il s'agit là d'une règle générale :

« L'offre de formation d'une université, quelle qu'elle soit, à un moment donné est le résultat d'un processus historique plus ou moins anarchique selon le mode de gouvernement de l'université, qui repose sur les stratégies des différents acteurs de l'université et qui est soumis aux exigences des différents partenaires (politique ministérielle, milieux politiques, socioprofessionnels, étudiants, etc.) et aux contraintes de l'environnement » (CNE, 2002b, p.60).

Dans son analyse des procédures d'habilitation des diplômes, Dejean (2002) montre bien comment même la rubrique « objectifs de formation » contient parfois un simple énoncé des contenus qui seront abordés dans la formation. Et, toujours d'après Dejean, un des mérites du développement des pratiques d'évaluation des enseignements résiderait précisément dans le fait qu'elles entraîneraient inévitablement un effort de formulation des objectifs de formation, puisque l'évaluation des enseignements doit se réaliser, selon l'arrêté qui la met en place, en relation avec les objectifs de ces enseignements.

Cette absence de définition explicite des compétences attendues en fin de formation se laisse aussi observer quand on analyse les problèmes rencontrés par les enseignants-chercheurs nouvellement recrutés. Le premier problème évoqué par ceux-ci, lors de la mise en place, dans une université française, d'une formation initiale qui leur était destinée, est le manque de définition des objectifs des cours qui leur sont confiés. Les enseignants-chercheurs disent manquer de repères sur le « produit attendu » : quelles sont les connaissances et/ou compétences à faire acquérir aux étudiants ? Les enseignants-chercheurs sont invités à assurer des cours, qu'ils n'ont parfois pas suivis eux-mêmes en tant qu'étudiants, sans qu'il soit précisé comment ces cours s'intègrent dans le programme global de la filière et quelles sont les finalités de cette filière. Ils éprouvent alors certaines difficultés à déterminer les exigences à attendre des étudiants et même les critères à partir desquels ils doivent déterminer ces exigences. C'est bien sûr lors de l'évaluation que ce problème se révèle le plus aigu : où placer la « barre » ? Quel est le niveau attendu d'un premier cycle ? En fonction de quels critères faut-il juger les étudiants ? Quel est le seuil de réussite ? Le problème concerne davantage les cours magistraux pour lesquels la « liberté académique » est vécue d'abord comme un isolement, voire un « abandon ». En ce qui concerne les TP et TD, leur prise en charge collective assure au moins, « sur le tas », une résolution en équipe de ces questions.

Sans doute devrait-on aussi se poser la question plus générale (mais qui déborde alors le cadre de cette mission) de savoir si l'absence d'explicitation des objectifs à atteindre au terme des formations universitaires n'est pas à mettre en relation avec la difficulté que rencontre actuellement l'université de masse à définir ses missions, surtout pour le premier cycle. En effet, identifier les objectifs explicites

²² La définition d'un programme peut aussi être le produit d'un moment, comme dans le cas, rapporté dans le cadre de la comparaison des deux DEUG d'histoire, d'un autre DEUG d'histoire qui, dès sa création, a pris une coloration fortement politique parce que les étudiants de cette université étaient très politisés.

des différentes formations exigerait de se positionner clairement au sein des débats qui animent actuellement le monde universitaire quant à ses missions et quant à la conception même de l'université. Le titre quelque peu embarrassé du dernier ouvrage de Renaut est représentatif de l'importance et de l'urgence de ces débats : « *Que faire des universités ?* ». Le relatif mutisme sur les connaissances et compétences qu'il s'agirait de faire acquérir aux étudiants est peut-être le reflet des tensions et des hésitations actuelles entre plusieurs modèles de l'Université. En ce que concerne l'épineuse question de la professionnalisation par exemple, Lessard et Bourdoncle (2002) montrent que le modèle libéral de l'université, qui prône une éducation globale, générale et pluridisciplinaire au sein de laquelle « *le savoir transmis et discuté est à lui-même sa propre fin* » (p. 136) répugne à introduire des savoirs utilitaires comme objectifs de formation. Le modèle humboldtien de l'Université de recherche suppose, quant à lui, que la meilleure préparation professionnelle reste la formation « *à et par la recherche* » : les principes qui guident l'activité scientifique (rigueur, objectivité, analyse, ...) sont aussi ceux que le professionnel doit apprendre à maîtriser pour progresser dans sa pratique. Dans le troisième modèle, on attend de l'université qu'elle soit au service de la société : les étudiants doivent être entraînés à utiliser leurs connaissances, de manière à répondre aux besoins sociaux. Comme le soulignent les auteurs, ces trois modèles continuent de nos jours à se disputer « *l'espace d'influence et de discours sur l'université* » (p.146).

Pour en revenir aux pratiques d'évaluation, si nous ne savons pas ce que les étudiants ont appris dans leur cursus et si l'évaluation se réalise sans trop expliciter les acquis attendus, ce serait parce que nous ne savons pas ce qu'ils auraient dû y apprendre et parce que l'université n'est pas au clair par rapport à ce qu'ils devraient y apprendre. L'explicitation des acquis attendus renvoie donc à la question fondamentale des finalités de la formation universitaire et notamment de son premier cycle : s'agit-il d'une formation propédeutique, par exemple en termes de maîtrise d'une méthodologie et d'élaboration d'un projet personnel ? D'une pré-professionnalisation ? Le premier cycle vise-t-il à développer une culture générale ? Une culture générale de sa discipline et de sa spécialité, comme le suggère Renaut (2002) ? ... Il nous a d'ailleurs été dit à plusieurs reprises lors des entretiens que « *l'on ne sait pas bien ce que doit acquérir tel ou tel type d'étudiants de premier cycle* » et que cette ignorance est explicable par le caractère très général des objectifs qui y sont poursuivis et par le fait qu'un consensus ne se dégage pas sur cette question.

À l'appui de cette hypothèse, on notera que le déficit d'explicitation des objectifs en termes de connaissances et de compétences à acquérir est davantage le fait du premier cycle que des cycles ultérieurs. On peut en effet comprendre que, si les maquettes des DESS et des nouvelles licences professionnelles contiennent des informations de ce type, ces dernières font davantage défaut pour les diplômes de premier cycle, dont les objectifs peuvent apparaître comme tellement généraux et génériques, comme la culture générale et la méthodologie de travail, qu'une explicitation détaillée des acquis attendus semble superflue et très délicate à réaliser.

3.3.2 Une absence de reconnaissance du caractère « professionnel » de l'acte d'évaluation

Malgré les réserves exprimées par les examinateurs externes interrogés par Warren Pipper (1994), les enseignants semblent faire preuve, dans la même enquête, d'une grande confiance dans leur jugement global intuitif et dans leur compréhension partagée et tacite des critères d'évaluation. Ils se disent convaincus que leurs jugements se rejoignent et marquent leur étonnement face aux études docimologiques qui identifient, au contraire, des problèmes de fidélité et de validité dans la passation des examens (cf. 3.4)²³. Deux tiers estiment, dans le même sens, qu'il y a peu de variation dans les niveaux des diplômes décernés par les différentes universités :

²³ Dans le même sens, Poirier (2001) note que si l'anonymat des copies n'est pas souvent garanti dans les faits, c'est notamment parce que les enseignants refusent de mettre en cause « l'honnêteté de leur évaluation ». Blais (1997) rapporte aussi que 89% des enseignants interrogés estiment qu'ils disposent des connaissances et des compétences docimologiques nécessaires à la bonne évaluation de leurs étudiants.

« On observe une propension des professeurs d'université à croire que des niveaux très élevés d'accord existent entre eux en tant qu'examineurs et à douter, dès lors, de l'efficacité, en regard de son coût, de l'application rigoureuse de procédures, par exemple, de correction multiple et anonyme. » (Warren Piper, 1994, p. 79-80)

L'auteur du rapport suggère l'interprétation suivante. Les professeurs d'université s'identifient davantage à leur discipline scientifique qu'à leur rôle d'enseignant ou d'évaluateur. En tant que membres d'une communauté scientifique, ils estiment partager une même idée de ce que représentent les critères de qualité et d'excellence au sein de leur discipline. Puisque les étudiants sont jugés sur ces critères qui constituent le ciment même de leur appartenance à une communauté de chercheurs, l'évaluation ne devrait dès lors pas soulever de problèmes docimologiques importants.

Une citation, extraite d'un entretien rapporté par Dejean (2002, p. 29), témoigne de cette confiance en un « collègue invisible » qui assurerait une cohérence aux critères d'évaluation des enseignants-chercheurs d'une même discipline : *« les experts savent quelque part ce qu'est un niveau de licence ou de maîtrise en droit ou en économie »*, ce qui justifie alors l'absence de définition d'objectifs de formation et donc de critères d'évaluation. Dans l'étude de cas rapportée en annexe 7.1, des enseignants rapportent qu'ils ont abandonné le recours à des grilles de correction parce que *« avec l'expérience, une pesée globale suffit ; avec l'habitude, on sait ce que vaut telle ou telle copie »*. Et ce que l'enseignant expérimenté cherche dans la copie, ce sont justement des signes de l'affiliation de l'étudiant à la communauté de recherche auquel il tente de s'affilier : *« un étudiant qui pose des questions sans trop y répondre de manière définitive, on sait que c'est quelqu'un qui a compris ce qu'est l'histoire »*.

En conclusion de son étude sur les pratiques d'évaluation dans les universités anglaises, Warren Piper (1994) regrette que l'évaluation ne soit pas considérée comme une des composantes de la « professionnalité » de l'enseignant-chercheur. Les enseignants du supérieur n'y sont pas formés. Ils apprennent « sur le tas », se définissent progressivement des règles pragmatiques et se rassurent en postulant une capacité à identifier le niveau « d'excellence », signe distinctif de l'université. Peu de réglementations explicites encadrent cet aspect de leur travail. Et quand elles existent, elles ne sont pas toujours connues et respectées. Dans l'enquête de Blais et al. (1997), 57% des enseignants interrogés estiment qu'aucune contrainte ne pèse sur leur façon d'évaluer les étudiants : ils se sentent donc libres de les évaluer comme bon leur semble. Mais, la même enquête montre qu'une part importante des enseignants interrogés contreviennent à une série de dispositions du règlement pédagogique de l'Université de Montréal. Ainsi, 62% des enseignants déclarent qu'il n'existe pas, dans leur unité, de jury d'examen, pourtant expressément prévu dans le règlement pour l'attribution définitive des notes. Le système littéral de communication des résultats (de A à E) ne semble pas adopté par les 60% des enseignants qui continuent à communiquer aux étudiants les résultats en pourcentage. La conclusion ne laisse pas d'inquiéter : *« Force est de constater que ceux qui ont répondu au questionnaire ou participé aux entrevues ne semblent pas se formaliser outre mesure de certaines dispositions du règlement pédagogique. »* (Blais et al, 1997, p.116).

Le manque de formation des enseignants-chercheurs à l'évaluation doit être replacé dans le contexte plus général de la quasi-absence de toute formation pédagogique du personnel enseignant du supérieur, même si des tentatives dans ce sens sont enregistrées çà et là et s'il ne faut pas négliger l'importance des dispositifs informels de formation, comme les réunions d'équipes d'enseignement (Donnay & Romainville, 1996). En France, les Centres d'Initiation à l'Enseignement Supérieur sont en charge d'une formation pédagogique initiale, mais tous les enseignants-chercheurs ne passent pas par les CIES et les programmes de formation offerts par les différents centres sont très divers (Bireaud, 1996). En particulier, l'évaluation des acquis ne se trouve pas au programme de tous les centres. Les premiers Services Universitaires de Pédagogie qui se mettent actuellement en place, comme à Lyon 1, offrent également des occasions de formation pédagogique, mais ils restent encore exceptionnels.

Si la pédagogie reste le parent pauvre de la formation de l'enseignant-chercheur, c'est notamment parce que les activités didactiques ne sont pas précisément celles qui sont le plus valorisées dans l'échelle implicite des valeurs universitaires. Une belle unanimité se dégage de l'ensemble des

rapports sur l'enseignement supérieur pour dénoncer le fait que, malgré la rhétorique flamboyante de l'interpénétration féconde des deux missions de l'université (l'enseignement et la recherche) comme élément constitutif de la formation universitaire, c'est essentiellement par la qualité de ses travaux de recherche qu'un enseignant-chercheur assure son recrutement d'abord, sa promotion et sa reconnaissance ensuite (Romainville, 1996). Le contexte actuel de forte concurrence internationale pour la publication des travaux de recherche n'est guère de nature à inciter les enseignants-chercheurs nouvellement recrutés à investir dans leur mission d'enseignement.

De plus, les activités didactiques et, en particulier, le choix de ce que chaque enseignant décide de mesurer lors de ses examens restent considérés comme relevant de la sphère privée (Dejean, 2002). L'enquête de Blais (1997) illustre aussi l'aspect individuel de l'activité d'évaluation : les enseignants disent montrer rarement leurs questions d'examen à leurs collègues. Ces derniers interviennent peu dans le choix des moyens d'évaluation et encore moins dans la correction : il est, par exemple, exceptionnel qu'un professeur sollicite l'avis d'un collègue en cas d'indécision quant à la note à attribuer à une copie.

Il faudrait enfin s'interroger sur l'impact de la conception qu'a l'enseignant du supérieur de son rôle en tant que formateur, sur la manière dont il envisage l'évaluation des acquis. Par exemple, on peut faire l'hypothèse qu'à la différence de ce que l'on observe dans l'enseignement obligatoire, le professeur peut se sentir essentiellement responsable de la qualité de ce qu'il fournit aux étudiants, mais pas des résultats de la formation, puisque la présence des étudiants résulte d'un choix libre et volontaire. La qualité de leurs acquis serait de leur ressort principal. Chercher à expliciter les acquis attendus supposerait, au contraire, que le formateur estime que sa responsabilité s'étend jusqu'aux effets de sa formation, ce qui le conduirait à s'interroger sur ce qui a été mis en œuvre pour garantir ces acquis.

3.3.3 Une évaluation à fonction sélective et non d'inventaire

En définitive, on pourrait risquer l'hypothèse selon laquelle l'évaluation des acquis est hétérogène et peu explicite quant à ses critères parce qu'elle remplit prioritairement des fonctions de sélection, de tri et de certification de réussite globale, ces fonctions ne nécessitant pas de disposer d'inventaires détaillés des compétences et connaissances acquises. Conçue comme une série d'obstacles permettant d'éliminer progressivement les étudiants les plus faibles, l'évaluation des étudiants, surtout ceux de premier cycle, ne chercherait pas prioritairement à valider des acquis précis. Comme le rappelle Hutmacher (2001), le diplôme universitaire possède une fonction de signal à l'égard du monde extérieur, lui indiquant que le détenteur de ce diplôme a réussi son passage par l'enseignement supérieur, ce qui garantit un certain niveau de compétences et l'acquisition de la culture du champ social correspondant. On se souviendra d'ailleurs que, dès 1964, Bourdieu et Passeron, dans leurs travaux sur les héritiers, soutenaient que l'absence d'explicitation des critères d'évaluation qui caractérise l'examen universitaire, « asile privilégié de l'irrationalité », privilégie les étudiants des classes cultivées « *puisque'ils détiennent, implicitement, le moyen d'y satisfaire* » (Bourdieu et Passeron, 1964, p. 113).

3.4 UN MANQUE DE VALIDITÉ, DE FIDÉLITÉ ET DE FIABILITÉ

Un quatrième thème de recherche sur les pratiques d'évaluation au supérieur a trait à la qualité docimologique des examens qui y ont cours. Certes, la critique docimologique n'est guère spécifique au supérieur. Cependant, l'importante hétérogénéité des pratiques (3.2) et l'absence d'explicitation des critères (3.3) posent, avec plus d'acuité encore qu'aux autres niveaux scolaires, la question de la fidélité, de la validité et de la fiabilité de l'évaluation.

L'analyse du fonctionnement des examens et de l'évaluation scolaire a une longue histoire. Au début des années 60, Piéron propose le terme de « docimologie » pour désigner cette nouvelle discipline, la science des examens. Essentiellement critique à l'origine, cette discipline a identifié une série de « dysfonctionnements » de l'évaluation scolaire (pour une synthèse de ces critiques, De Landsheere, 1980). Elle s'est ensuite orientée vers l'étude du processus de jugement des notateurs (Noizet et Caverni, 1978).

Pour l'essentiel, la docimologie a montré que les examens scolaires souffraient de deux défauts majeurs. D'une part, leur *fidélité* n'est pas toujours assurée : des études empiriques, voire expérimentales ont montré clairement que l'évaluation, quelle que soit la discipline, n'aboutit pas à des résultats similaires si elle est reproduite à un autre moment ou si elle est réalisée par une autre personne.

- La liberté laissée aux enseignants dans un système d'évaluation interne, endogène à l'établissement, conduit à des injustices : les enseignants adaptent en effet leurs examens aux caractéristiques de leurs élèves avec, comme résultat, que certains d'entre eux réussissent dans une classe alors qu'ils auraient échoué dans une autre, d'un niveau moyen plus élevé.
- Pour un même travail, les désaccords entre correcteurs sont importants en termes de points, de moyenne, de dispersion et de décision (de réussite ou d'échec).
- Un même correcteur n'est pas égal à lui-même.

D'autre part, la *validité* de l'évaluation a plus récemment fait l'objet de critiques : l'évaluation devrait mesurer ce qu'elle prétend mesurer et rien que ce qu'elle prétend mesurer. Une évaluation valide s'en tient donc à chercher à savoir si les objectifs de l'enseignement ont été atteints par les élèves et les épreuves ne devraient mesurer que cela et le mesurer de manière représentative. Or, les examens mesurent parfois des compétences qui n'ont pas fait l'objet d'un enseignement et, de plus, le correcteur établit sa note sur la base d'autres critères que la seule performance scolaire de l'élève. Citons notamment quelques biais particulièrement bien étudiés :

- la contagion des évaluations : le correcteur se laisse influencer par les évaluations antérieures. On explique cet effet par la recherche d'une consonance cognitive dans le jugement porté sur les individus ;
- l'effet de halo : la notation peut être influencée par la connaissance que peut avoir l'examineur du niveau scolaire de l'élève, de son origine socioculturelle et ethnique ;
- l'effet d'ordre et de position de correction.

Ce qui est moins connu, c'est que des études de docimologie, parfois anciennes, ont aussi été réalisées dans l'enseignement universitaire, avec des résultats similaires (Ashcorft & Palacio, 1996 ; Beattie, 1995 ; Brown & Glasner, 1999 ; Heywood, 2000 ; Warren Piper, 1994).

Ainsi, dès les années 30, une série d'études ont été consacrées à la **fidélité** des examens universitaires : des correcteurs ont été invités à noter à nouveau des copies qu'ils avaient déjà corrigées. Les notations d'une centaine de professeurs de mathématiques ont été comparées à propos de cinq mêmes copies. Les résultats sont sans appel : dans la détermination de la note, la part du correcteur se révèle finalement plus décisive que celle de la performance réalisée par l'étudiant ! Ces études ont été confirmées par un programme de recherche de la Society for Research into Higher Education réalisé dans les années 60, puis par des études portant plus spécifiquement sur des disciplines particulières, notamment les sciences médicales et les sciences de l'ingénieur. Dans l'étude de Warren Piper (1994), les examinateurs externes s'interrogent aussi régulièrement sur la fidélité de l'évaluation. Ils se déclarent parfois surpris de la diversité des notes en cas d'examineurs multiples. Ils s'interrogent aussi sur la comparabilité des jugements évaluatifs dans le temps et entre les filières : comment garantir qu'un diplôme sanctionne un même niveau d'études d'année en année et que des standards de qualité aient la même signification au sein de différentes filières ? Comment s'assurer, par exemple, qu'une mention corresponde au même niveau d'excellence dans des disciplines différentes ?

Une des explications du manque de fidélité réside dans le fait que les critères d'évaluation varient considérablement d'un examinateur à l'autre, parce que ces examinateurs sont guidés par des cadres de référence différents pour juger leurs étudiants (Warren Piper, 1994 ; Webster et al., 2000). Par exemple, certains correcteurs tiennent compte de la performance des autres étudiants, d'autres s'attachent à repérer les signes d'un progrès depuis le début du cours. Le jugement porté sur une dissertation est dépendant de ce que le correcteur valorise dans un tel travail : certains privilégient

l'apport de données empiriques, d'autres l'aspect théorique. Pour une même filière, les jurys de différents établissements n'attribuent pas les mentions selon les mêmes critères et sont ignorants des critères appliqués ailleurs. Par exemple, la mention la plus élevée est accordée par certains examinateurs si le candidat se distingue par son originalité, par d'autres si le candidat fait montre de compétences de communication remarquables. Cette observation est à mettre en parallèle avec le fait que peu de discussion et d'explicitation semble être consacré, au sein des unités de formation, à l'établissement collectif des critères d'évaluation.

Même quand les correcteurs se réfèrent à des critères identiques, ils ne les font pas intervenir selon la même pondération²⁴ et ils ne leur attribuent pas la même signification. Ainsi, si les enseignants sont assez unanimes pour identifier les critères pertinents pour classer les étudiants (l'analyse critique, par exemple), on observe que des comportements précis très différents sont retenus comme indicateurs de ce critère selon la discipline enseignée et même parfois selon les théories privilégiées au sein de la même discipline (Ecclestone, 2001). Dans le même sens, Webster (2000) a demandé à 24 enseignants de préciser ce qu'ils entendaient quand ils annonçaient à leurs étudiants qu'un de leurs critères d'évaluation résiderait dans la « qualité de l'analyse ». Pour un premier groupe d'enseignants, l'analyse est définie comme la décomposition d'un tout en ses différentes parties, de manière à mieux comprendre les relations qui les unissent. Pour un second groupe, il s'agit de l'application d'outils méthodologiques aux données recueillies. Neuf autres des 24 enseignants interrogés fournissent chacun une définition qui ne se rapproche ni de celle du premier groupe ni de celle du second. La même étude révèle que les critères effectivement pris en compte lors de la correction diffèrent parfois considérablement des critères annoncés. Ainsi, tous les critères annoncés ne sont pas nécessairement appliqués et certains critères effectifs, induits à partir des annotations des copies, n'avaient pas été annoncés. Dans le même sens, Dejean et Magoga (2001) ont observé que les qualités strictement linguistiques des réponses écrites des étudiants (de l'orthographe à la construction discursive) interviennent dans la notation à des degrés très divers selon le correcteur.

Un problème de fidélité qui a fait couler beaucoup d'encre a trait à la question de savoir si l'université, à la suite de sa massification, n'a pas eu tendance à baisser ses **normes d'excellence**. On s'en doute, la question est épineuse et sensible. Dans ce cadre, on a constaté un phénomène étrange au Royaume-Uni, à savoir l'augmentation considérable du niveau des notes de sortie, dans le même temps que les ressources financières par étudiant chutaient drastiquement. Ainsi, les mentions de première et de deuxième classes supérieures sont passées de 29,3% en 1969 à 47,7% en 1989 (Gibbs & Lucas, 1997). Comme nous l'avons vu au point 3.2, il semble que l'augmentation de la part du contrôle continu soit en grande partie responsable de cette évolution. Mais alors, se pose la question de savoir si ces mentions conservent toujours la même signification. En effet, on ne mesure pas la même chose par un examen final à livre fermé et à temps limité que par des travaux personnels continus : l'étudiant peut y passer le temps qu'il souhaite, il peut consulter des documents, il ne se fie pas seulement à sa mémoire, il peut même mettre à contribution d'autres personnes, ...

Plusieurs études récentes montrent que des problèmes de **fidélité** se rencontrent aussi pour l'obtention du diplôme de **doctorat**. Johnston (1997) a ainsi réalisé une analyse détaillée de rapports de thèse. Sur les 16 thèses analysées et jugées chacune par quatre professeurs, trois d'entre elles font l'objet de désaccord entre les enseignants à propos de la décision d'échec ou de réussite. Pour l'une des thèses analysées, un rapport relève la qualité et l'originalité de la revue de la littérature (« éclectique, organisée et cohérente »), alors que le rapport de l'expert prônant l'échec contient sept pages de critiques sur cette partie de la thèse, qui se terminent en regrettant que le candidat n'ait pas eu recours à la structure classique et académique d'un relevé de la littérature. Dans un autre cas, c'est manifestement une incompatibilité idéologique qui explique le désaccord. Des commentaires identiques n'ont, de plus, pas le même impact chez tous les rapporteurs : des suggestions de modifications de forme (écriture, références, ...) peuvent justifier un avis d'acceptation sous condition

²⁴ Ce qui explique que des commentaires identiques ou proches apparaissent sur des copies ayant obtenu des notes très différentes ou que des notes identiques soient justifiées par des commentaires antinomiques.

de modification, tout comme elles peuvent être simplement mentionnées après un avis favorable sans condition. En ce qui concerne le type de critères utilisés pour juger de la qualité de la thèse, les rapporteurs se réfèrent tous à des critères de qualité formelle (écriture et présentation), mais divergent quant aux autres critères (gestion de la complexité, originalité, pensée critique, produit publiable, ...), au point que peu de thèmes sont communs à l'ensemble des rapports. Enfin, un même critère peut revêtir des sens différents : ainsi, quand les rapporteurs se réfèrent au caractère « publiable » du travail de recherche, cela signifie, pour certains, que la thèse est publiable telle quelle ; pour d'autres, que le travail mérite publication, mais après avoir subi une série de révisions parfois importantes. Dans le même sens, Morley et ses collaborateurs (2002) ont décrit l'importante hétérogénéité des pratiques d'évaluation de la thèse de doctorat selon les institutions, les facultés et les départements : poids relatif de l'oral par rapport à l'écrit, modalités de rapports de thèse et de constitution du jury.

La **validité** des examens universitaires a aussi fait l'objet de questionnements critiques. On s'est ainsi interrogé sur la *validité prédictive* de ces examens : permettent-ils de prédire les performances ultérieures, voire la réussite professionnelle (Goldschmidt, 1992) ? On a aussi mis en évidence l'existence de *biais idéologiques* dans la correction de copies (Husbands, 1976). Cette étude a porté sur un département de sciences sociales au sein duquel deux écoles de pensée coexistaient, l'une centrée sur l'action sociale et l'autre sur des approches sociales cliniques. Six enseignants, dont on connaissait le « penchant » épistémologique, ont été invités à noter des copies d'étudiants, la moitié de ces copies manifestant une certaine sympathie pour la première école de pensée et l'autre moitié pour la seconde. L'existence de biais idéologiques n'est pas systématique : certains correcteurs semblent faire fi de leurs options méthodologiques personnelles. Mais d'autres surévaluent assez systématiquement les copies qui valorisent les cadres d'analyse qu'ils privilégient eux-mêmes en tant que chercheurs. L'auteur conclut son étude en montrant comment l'attitude stratégique de certains étudiants qui adaptent leur discours aux positions épistémologiques qu'ils pensent être celles de leur correcteur est finalement assez rationnelle, alors qu'elle est parfois sévèrement punie, notamment par les évaluateurs externes qui y voient une sorte de jeu de séduction inutile et peu « universitaire ».

L'existence de *biais culturels* dans les procédures d'évaluation des étudiants a aussi fait l'objet d'investigation, dans le contexte d'un enseignement supérieur mondialisé. On s'est notamment interrogé sur l'adéquation de certaines formes d'examen aux différents publics d'étudiants, en particulier les étudiants étrangers et ceux qui appartiennent aux minorités ethniques ou culturelles. Ainsi, De Vita (2002) a montré que les différentes formes d'examen n'offrent pas aux étudiants d'origine culturelle différente les mêmes opportunités de manifester leur maîtrise du cours. Par exemple, les étudiants d'origine chinoise inscrits dans les universités anglaises semblent, par certains côtés, désavantagés par les QCM car ils répugnent à choisir au hasard une réponse en cas de doute. Par d'autres aspects cependant, les examens écrits à réponse longue les désavantagent également, comme tous les étudiants non anglophones d'ailleurs, tant ce type d'examen mesure autant (sinon davantage) les compétences langagières que la stricte maîtrise des connaissances.

Pour l'anecdote, d'autres biais, parfois beaucoup plus pragmatiques, ont aussi fait l'objet d'études docimologiques. Ainsi, les correcteurs de travaux d'étudiants auraient tendance à se laisser influencer par la qualité de présentation du travail, y compris par des caractéristiques typographiques et de mises en page de ceux-ci. Une taille de caractère 12 avec un espacement simple de 1,5 semble assurer les meilleurs résultats... (Kangis, 2001). Autre exemple, des copies d'examens dont la calligraphie est déficiente sont assez systématiquement attribuées à des étudiants plutôt qu'à des étudiantes (Brown & Glasner, 1999).

Une autre observation interroge la validité de l'évaluation des acquis. Comme nous l'avons déjà signalé, l'évaluation se réalise parfois sur la base de travaux réalisés par les étudiants en cours d'année. Or, un certain nombre d'études rapportent que les notes des examens terminaux ne sont corrélées que modérément avec les notes remises pour ces travaux de contrôle continu (Yorke et al., 2000), ce qui indiquerait que ces deux formes d'évaluation mesurent des compétences distinctes, alors que leur importance relative semble tributaire de la filière et surtout de la taille des groupes.

En Allemagne, un débat important a vu le jour, au milieu des années 70, sur la validité de l'examen oral, au point de réclamer son abolition. Au-delà des problèmes docimologiques classiques, il était reproché à l'oral, dans l'élan des mouvements contestataires étudiants, d'être le lieu privilégié de l'imposition de normes arbitraires. La situation de l'oral cristallisait la relation inégale de pouvoir entre un enseignant, imposant ses vues et ses normes, et un étudiant qui n'a qu'à s'y soumettre. De plus, les règles du jeu qui régissent implicitement cette rencontre particulière n'étaient pas également maîtrisées par l'ensemble des étudiants, en particulier ceux qui étaient issus des classes sociales défavorisées (Khem, 2001).

Enfin, des interrogations critiques portent aussi sur la **fiabilité** de l'évaluation des acquis, c'est-à-dire sur la capacité qu'a l'université de masse à assurer que les examens se déroulent sans incidents. À l'occasion de plusieurs entretiens, des doutes ont été émis à propos des garanties qu'offre l'université quant à la fiabilité des procédures et des dispositifs d'évaluation, surtout dans les premiers cycles massifiés. Des problèmes particuliers ont été évoqués : les mauvaises conditions de passation des épreuves écrites qui aboutissent parfois à l'impossibilité de s'assurer du caractère individuel de l'examen ; la répétition des mêmes questions d'année en année ; la présence aux examens d'étudiants qui ont disposé à l'avance des questions ; les fraudes en cours d'examen, notamment à l'aide d'outils technologiques de plus en plus pointus.

Comme pour confirmer ces interrogations critiques, on trouve dans la littérature contemporaine à la massification une série d'articles sur la fraude aux examens (Franklyn-Stokes et Newstead, 1995 ; Wankat et Oreovicz, 2002). Les comportements frauduleux prennent des formes très variées dans le supérieur. Certains, parmi les moins graves, touchent un étudiant sur deux dans le premier cycle : permettre à un autre étudiant de copier un devoir ou une préparation ; paraphraser sans citer explicitement ; modifier, voire inventer des données ; surévaluer le travail d'un pair lors de l'évaluation par les pairs ; recopier un travail antérieur ; inventer des références. Les comportements de fraudes durant les examens seraient plus rares, bien qu'ils puissent toucher plus d'un étudiant sur dix : utiliser du matériel non autorisé, passer un examen à la place de quelqu'un d'autre, obtenir à l'avance des informations sur l'examen, copier sur un voisin. En tout cas, la fraude serait plus présente que ne le pensent les enseignants-chercheurs et se développerait corrélativement à l'évolution actuelle des procédures d'évaluation. En effet, pressée d'évaluer, plus souvent suite à la modularisation et à la semestrialisation, un nombre d'étudiants plus important, l'université de masse se tournerait davantage vers des formules de questionnaires à choix multiple et à correction automatique, forme d'examen qui semble la plus propice à la fraude. Par ailleurs, les comportements frauduleux sont moins fréquents quand l'évaluation a la réputation d'être correctement et justement conduite et quand elle vise à mesurer des compétences qui dépassent la restitution, par exemple via des examens à livre ouvert.

Enfin, les études docimologiques réalisées dans l'enseignement supérieur dénoncent aussi, comme aux autres niveaux scolaires, l'usage intempestif des nombres dans les procédures d'évaluation et surtout le recours à certaines opérations arithmétiques (de moyenne notamment) réalisées sur ces nombres, alors que la nature qualitative des variables mesurées ne l'admet, en principe, pas. Dalziel (1998), par exemple, a montré comment la manière d'agrèger les différentes notes des étudiants aux examens modifie considérablement la répartition des mentions au sein d'un groupe d'étudiants.

En conclusion de son relevé de la littérature sur la docimologie dans le supérieur, Beattie (1995, p. 316) note que *« l'impression qui se dégage des diverses études est que les méthodes d'évaluation généralement utilisées pour s'assurer que des connaissances ont bien été acquises ont grand besoin d'être examinées. Elles n'ont guère de chance, sous leur forme actuelle, d'être considérées comme acceptables par ceux qui militent en faveur d'une approche plus rigoureuse de l'évaluation des résultats des études »*.

3.5 UNE ÉVALUATION QUI « PILOTE » LES ÉTUDIANTS

Le domaine sans doute le plus documenté concernant les pratiques d'évaluation dans le supérieur a trait aux effets de ces pratiques sur l'apprentissage des étudiants. Depuis longtemps en effet, des recherches ont été consacrées à **l'impact des pratiques d'évaluation sur la qualité des**

apprentissages réalisés par les étudiants. Les résultats de ces recherches convergent vers un résultat principal : les étudiants sont véritablement « pilotés » par les exigences de l'évaluation. Ils sont à l'affût de ce qui leur sera demandé aux examens et ont tendance à y adapter leurs manières d'étudier (Biggs, 1999 ; Romainville, 2000). Si l'évaluation apparaît aux enseignants comme la toute fin d'un processus, elle est, par contre, le début de toute chose pour les étudiants.

3.5.1 L'évaluation par la restitution appelle l'étude en surface

Les méthodes d'évaluation ont parfois tendance, au premier cycle surtout, à privilégier des mesures de restitution des matières enseignées, notamment parce que ces méthodes sont les plus aisées à mettre en œuvre au sein des grands groupes de l'université de masse (cf. 3.1). On constate en effet que bon nombre de questions d'examen n'exigent pas beaucoup plus qu'une reproduction de parties de photocopié. Ces pratiques d'évaluation incitent alors les étudiants à privilégier une étude en surface des matières. Plusieurs observations vont dans ce sens. Des questions portant sur la restitution de faits pointus engendrent une étude mécanique, « par cœur », morcelée et superficielle (Montgomery, 1995). Une accumulation d'évaluations sommatives partielles encourage l'étude par la seule mémorisation (Tan, 1992). L'opacité des méthodes d'évaluation est aussi responsable du développement de l'approche en surface : l'étudiant qui ne sait pas très bien ce qui lui sera demandé à l'examen se réfugie dans une reproduction stricte des contenus du cours (Edwards & Knight, 1995).

À l'inverse, les étudiants adoptent plus volontiers une approche en profondeur quand l'évaluation se réalise par une production écrite ouverte plutôt que par des examens à questions fermées, le questionnaire à choix multiples en particulier (Fallows & Chandramohan, 2001). D'ailleurs, les étudiants qui réussissent aux QCM sont ceux qui ont bien compris que cette forme d'évaluation mesure la maîtrise de connaissances ponctuelles (faits, dates, formules, ...) : ils s'y préparent en conséquence en négligeant d'exercer des opérations intellectuelles plus englobantes, comme la comparaison (Scouller, 1998).

En adaptant leurs stratégies à l'évaluation, les étudiants réalisent en définitive des apprentissages de nature différente. Ainsi, les productions écrites d'étudiants engagés dans une pédagogie active manifestent davantage de traces d'opérations intellectuelles de haut niveau (comparaison, généralisation, classification, ...) que les écrits d'étudiants qui ont assisté, pour la même matière, à un cours magistral suivi d'un examen traditionnel (Tynjälä, 1998). Dans le même sens, l'introduction d'évaluations formatives intermédiaires permet aux étudiants de réaliser des apprentissages de qualité supérieure (Greer, 2001). Bref, quand on déplore que les étudiants ne travaillent qu'en vue de la réussite d'un test, c'est surtout la mauvaise qualité de ce test qui est en cause : s'ils étudient superficiellement, c'est qu'ils savent qu'ils seront interrogés superficiellement²⁵.

Cet impact des procédures d'évaluation sur les apprentissages étudiants est aussi évoqué dans certains rapports du CNE. Ainsi, le rapport transversal portant sur les formations de pharmacie signale que « *dans trop de cas encore, des examens multiples et parcellaires sont imposés par les professeurs, dont chacun vérifie que son cours est appris, sans réflexion d'ensemble sur ce qui doit être appris, et non enseigné (toutes les classes de médicaments), et sur le type d'entraînement qui favoriserait l'acquisition de concepts et de mécanismes intellectuels rigoureux. Trop souvent, les examens confortent une démarche encyclopédiste et morcellent davantage encore les connaissances* » (CNE, 1998a, p. 39). L'évaluation des différentes UFR se réfère également à ce critère : il est régulièrement reproché à telle ou telle UFR de recourir à des formes d'examen peu appropriées à développer les compétences de haut niveau attendues, comme l'esprit de synthèse. De l'avis des étudiants, « *pas de temps, pas de place à la réflexion : il faut redonner le cours mot à mot et de préférence avec les virgules, ce qui fait qu'on ne retient pas* » (CNE, 1998b, p. 62). C'est notamment au nom de leur

²⁵ Même s'il semble, heureusement, exister une certaine immunité des meilleurs étudiants aux effets des mauvaises pratiques d'évaluation des acquis : les meilleurs étudiants en médecine, même s'ils sont confrontés à des tests qui n'exigent que de la mémorisation, continuent à étudier dans une perspective principale de compréhension et d'application de leurs connaissances (LindBlom-Ylänne & Lonka, 2001).

impact positif sur la qualité des apprentissages des étudiants que le CNE recommande que des UFR accordent plus de place aux oraux et aux questions ouvertes.

3.5.2 L'évaluation par la restitution est peu compatible avec les plus hautes finalités de l'enseignement universitaire

Ce questionnement critique par rapport aux effets délétères des pratiques d'évaluation sur l'apprentissage des étudiants s'est exacerbé quand l'enseignement universitaire a été prié de participer plus directement au développement économique et social des pays et de « rendre compte » de l'utilisation de la part des finances publiques qui lui était consacrée (« accountability »)²⁶. Les critères retenus par l'Université pour évaluer les étudiants semblaient fort éloignés de ceux qui ont cours dans les environnements de travail actuels pour évaluer l'efficacité professionnelle. La remise en cause des pratiques d'évaluation est ici significative d'une volonté d'assigner de nouveaux objectifs à l'enseignement tertiaire et de créer des synergies plus étroites entre le système éducatif et les politiques économique et sociale de chaque pays (Neave, 1996). Par exemple, Edwards et Knight (1995) déplorent que l'évaluation académique des acquis privilégie les épreuves qui portent sur les connaissances déclaratives alors que le monde du travail souhaite des étudiants qui maîtrisent des connaissances procédurales, mobilisables pour l'action. Dans la même veine, ils critiquent le recours trop exclusif à l'essai personnel, alors que ce serait les compétences de communication orale et de travail de groupe qui seraient prioritairement recherchées par les employeurs. Au total, les compétences les plus estimées par les employeurs, comme la capacité de résoudre des problèmes réels, ne seraient acquises par les étudiants que fortuitement, comme des conséquences « collatérales » d'un curriculum universitaire formel qui ne s'en préoccupe pas directement (Montgomery, 1995).

Dans le même sens, Beattie (1995) estime que les pratiques actuelles d'évaluation restent avant tout conçues pour tester l'acquisition de connaissances factuelles. Les détracteurs de l'Université ont alors beau jeu de prétendre que cette dernière remplit la tête des étudiants de faits arides et sans intérêt qui, même s'ils sont retenus pour les examens, sont ensuite rapidement oubliés. De plus, ce qui est privilégié dans les pratiques d'évaluation reste la maîtrise de savoirs déclaratifs formels²⁷ et les problèmes issus de situations pratiques ne sont considérés qu'à titre d'exemples. Les étudiants ne sont incités à établir des liens entre ces connaissances formelles apprises, ni avec leur savoir naturel, ni avec la résolution de problèmes réels. Ils développent dès lors des « savoirs morts », c'est-à-dire des connaissances déconnectées, apprises par cœur, peu mobilisables dans l'action. Ramsden (1988) décrit ainsi comment des étudiants qui ont pourtant réussi des examens universitaires de connaissance n'ont pas modifié durablement leur manière d'appréhender les phénomènes de la vie quotidienne en rapport avec leur discipline de base. Or, ce sont surtout des « savoirs vivants », des compétences, des savoirs pour penser et pour agir que l'ensemble du système éducatif, du primaire au supérieur, est actuellement invité à développer chez les jeunes, sous la pression du monde économique et social.

Enfin, comme l'ensemble du processus évaluatif est aux mains de l'enseignant (de la définition des objectifs du cours jusqu'à la détermination des critères), les pratiques actuelles ne semblent pas en mesure de faire acquérir aux étudiants certaines compétences qui sont pourtant au cœur même des finalités de l'enseignement universitaire, comme l'autonomie, l'esprit d'initiative, la responsabilité et l'esprit critique. Il conviendrait dès lors de modifier les pratiques d'évaluation (cf. 5.2.4), de manière à maintenir la poursuite d'objectifs de haut niveau, comme les capacités de réflexion et de synthèse, par

²⁶ C'est sans doute au Royaume-Uni que cette tendance a été la plus marquée. En 1990, le Ministère de l'emploi s'est ainsi immiscé dans ce qui apparaissait jusque-là comme du domaine privé des universités, en réclamant un enseignement supérieur « *plus souple, plus adapté et plus sensible à l'évolution des besoins du marché du travail et de la vie active* » et donc « *d'avantage axé sur ce que les étudiants doivent apprendre et sur la manière dont ces connaissances peuvent être évaluées* » (Beattie, 1995, p. 308).

²⁷ La question emblématique, certes caricaturale mais bien réelle, de cette tendance est le célèbre « Que savez-vous de... »...

exemple en proposant des épreuves synthétiques portant sur plusieurs enseignements (Boud, 1990 ; Poirier, 2001 ; Reynolds & Trehan, 2000).

4 POURQUOI DEVRAIT-ON EN SAVOIR DAVANTAGE ? LIMITES ET LACUNES DES PRATIQUES ACTUELLES D'ÉVALUATION DES ACQUIS

Si l'absence de standardisation et d'explicitation des pratiques d'évaluation des acquis semble appartenir à une longue tradition universitaire, les propositions d'amélioration de cette situation doivent impérativement reposer sur un argumentaire solide, si elles veulent avoir quelque chance d'être mises en œuvre. L'élaboration de cet argumentaire, auquel la quatrième partie de ce rapport est consacrée, s'est réalisée en deux temps. Dans un premier temps, les plus récents rapports prospectifs sur l'enseignement supérieur ont été consultés, en cherchant à savoir si l'amélioration de la qualité de l'évaluation des acquis faisait partie des enjeux majeurs identifiés par ces rapports. Dans un second temps, une liste des effets dommageables de la situation actuelle a été dressée. Cette liste définit, en creux, les raisons qu'il y aurait à prôner certains changements.

4.1 LES RAPPORTS PROSPECTIFS SUR L'ENSEIGNEMENT SUPÉRIEUR

Autour des années 90, l'Université a été confrontée à une série de transformations majeures, qui ont profondément remis en cause ses modes traditionnels de fonctionnement. Il suffit de penser à l'explosion des effectifs étudiants, d'une part, et au mouvement de la « responsabilité » et de la « qualité », d'autre part. Les rapports prospectifs se sont multipliés au cours de cette décennie, qu'ils soient **nationaux** (Commission Attali (1998) en France, Commission Boyer (1998) aux USA, Rapport Dearing (1997) au Royaume-Uni, Rapport du Conseil supérieur de l'éducation du Québec (1995) ou **internationaux** (Conférence des recteurs européens, 1997; Conférence mondiale de l'UNESCO sur l'enseignement supérieur, 1998 ; Conseil de l'Europe, 1998 et OCDE, 1998). En particulier, de nombreux documents ont été produits au niveau européen, dans la perspective de la création d'un espace européen de l'enseignement supérieur, avant et après la Déclaration de Bologne (Confédération des conférences des recteurs de l'UE, 1999).

La question de l'évaluation des acquis des étudiants ne semble pas occuper a priori une place prépondérante au sein de ce que ces rapports identifient comme les défis majeurs de l'enseignement tertiaire. À l'appui de cette conclusion, on peut mentionner que les rapports du CNE produits durant cette même décennie, qu'ils portent sur l'évaluation institutionnelle des établissements, sur les disciplines ou sur les problèmes généraux de l'enseignement supérieur (CNE, 1995, 1997 et 1999), accordent peu de place à cette question²⁸.

À y regarder de plus près cependant, la question de l'évaluation des acquis est en fait abordée indirectement dans ces rapports prospectifs, selon deux points de vue : la définition de standards de qualité sur lesquels pourrait se fonder la confiance dans les acquis développés par les étudiants et la question de l'employabilité des diplômés.

4.1.1 La certification de la qualité des acquis

Certains de ces rapports traitent de la question de l'évaluation des acquis sous l'angle de la nécessité de maintenir des standards de qualité dans ces acquis. C'est particulièrement net dans le **Rapport Dearing**. Créée en 1996 au Royaume-Uni, la Commission Dearing a été mise en place à la suite d'une menace de la Commission des vice-présidents d'université et des chefs d'établissements (CVCP) d'exiger des droits de scolarité élevés aux étudiants, pour compenser la chute importante des crédits accordés au supérieur, en particulier la baisse spectaculaire du financement par étudiant qu'avait entraînée l'explosion des effectifs de 1989 à 1995 (+ 50%) (Shattock, 1999). L'effacement de la ligne de démarcation entre les universités et les « polytechnics » n'avait fait qu'aggraver la

²⁸ Il faut cependant signaler qu'une des évaluations les plus récentes, celle qui porte sur les formations juridiques de base, se penche davantage sur ce problème. Une part non négligeable du questionnaire d'enquête est consacrée à l'organisation des examens. Elle comporte même un nombre de questions plus important que la partie consacrée à l'organisation pédagogique.

suspicion d'une chute de la qualité de la formation dispensée dans le supérieur. La Commission Dearing aborde la question de l'évaluation des acquis en termes de procédures à mettre en place ou à développer, comme les examinateurs externes, pour assurer une plus grande uniformité de la qualité des diplômes britanniques et pour offrir des garanties d'une constance de niveau de ces diplômes octroyés de manière autonome par les différents établissements.

L'uniformité et la constance de la qualité des acquis ne peuvent être assurées, selon le Rapport Dearing, que par une plus grande transparence de ces acquis. En effet, comme l'ensemble de procédures d'examens est aux mains des communautés locales d'enseignants-chercheurs (de la définition des objectifs au choix des outils de mesure), la société ne sait pas sur quels critères les étudiants sont jugés et finit par douter de l'existence de normes prédéfinies, en soupçonnant le supérieur de s'adapter aux pressions sociales. À titre d'exemple, le *Times* s'était inquiété de l'inflation des mentions de première classe et faisait remarquer ironiquement qu'il était difficile à croire qu'elle résultait d'une amélioration de la qualité de l'enseignement ou de la croissance des talents des étudiants... La Commission souhaite dès lors que soient fixés des critères nationaux définissant des niveaux minimums requis pour chaque diplôme. En cas de non-respect de ces critères par un établissement, une Agence pour la qualité pourrait recommander la suppression du financement de cet établissement pour ce diplôme particulier.

La standardisation et l'explicitation des pratiques d'évaluation sont donc résolument envisagées dans la perspective du maintien d'une certaine qualité d'un enseignement massifié et désargenté. Cette préoccupation de nos voisins anglais a abouti à la création de l'importante Agence pour la qualité (Quality Assurance Agency for Higher Education²⁹). L'évaluation par compétences est régulièrement désignée comme une des pistes possibles pour résoudre les problèmes évoqués ci-dessus.

4.1.2 L'employabilité

Depuis la Déclaration de la Sorbonne qui a initié le mouvement vers un espace européen de l'enseignement supérieur, la question de l'adéquation des connaissances et des compétences acquises par les étudiants par rapport au marché de l'emploi a été posée avec plus d'acuité (Haug, 2001). Néologisme peu élégant, l'employabilité est désormais présentée comme une des finalités de l'enseignement universitaire, à côté de ses missions traditionnelles de recherche et d'enseignement. Cette nouvelle préoccupation exige que soient mieux connus les acquis des étudiants à la sortie des études, de manière à pouvoir discuter de leur pertinence par rapport aux exigences du monde économique et social. Ce deuxième point de vue à partir duquel la question de l'évaluation des acquis est abordée dans les rapports prospectifs touche donc aux efforts d'explicitation qui devraient être entrepris en ce qui concerne les compétences et les connaissances acquises par les étudiants en fin de formation, de manière à vérifier qu'elles soient bien de nature à leur assurer une insertion sur le marché de l'emploi.

Il n'est ici pas question de rentrer dans le débat hautement sensible de la professionnalisation des études universitaires, tant ce nouveau défi complexe adressé aux universités est loin d'être exempt de questions non encore résolues (Hutmacher, 2001 ; Mignot-Gérard et Musselin, 2002 ; Renaut, 2002).

- Privilégie-t-on le paradigme humaniste (« former l'homme à être un homme de manière à ce qu'il sache occuper tout poste ») ou des exigences de professionnalisation étroite ?
- Est-il vraiment possible de définir déductivement le contenu disciplinaire des formations à partir d'une analyse fine des activités professionnelles ?
- L'imprécision, voire la contradiction ne président-elles pas aux demandes exprimées par le monde du travail, par exemple quant au type et à la pondération des savoirs et savoir-faire nécessaires à l'exercice de telle ou telle profession et quant à la part des compétences de savoir-être et d'attitudes ?

²⁹ Cette agence gère notamment un site Internet dont une part est consacrée à l'amélioration des procédures d'évaluation des acquis (<http://www.qaa.ac.uk/>)

- Comment concilier professionnalisation des études et raccourcissement constant de la durée de vie des savoirs ?
- Quelle est la validité de l'anticipation des besoins futurs, de la demande du marché du travail et de son évolution ?

Sans entrer dans ce débat complexe, notre propos est de souligner que la rencontre de cet enjeu de professionnalisation des études universitaires passe par un nouveau mode de définition des contenus et des parcours de formation (Hutmacher, 2001). Cet enjeu contribue donc à prôner une définition des formations en termes de connaissances et de compétences attendues, qu'elles soient pointues ou générales (selon la conception que l'on a de la professionnalisation) et une modification substantielle des pratiques d'évaluation.

La diversification du type de carrières vers lesquelles se dirigent les étudiants complexifie encore la situation. Alors que le secteur public était traditionnellement le principal employeur des diplômés universitaires, le secteur privé, dans une économie de la connaissance, a tendance à en engager aussi de plus en plus, en valorisant une formation solide et ouverte qui constituerait, à ses yeux, le gage d'une capacité d'adaptation à des situations nouvelles et imprévisibles (Hutmacher, 2001). Se pose alors la question de la définition des acquis à installer chez des étudiants à l'avenir professionnel incertain et diversifié. Cette question a d'ailleurs été abordée à l'occasion de l'étude de cas portant sur deux DEUG d'histoire (cf. 7.1). Alors que traditionnellement, la filière histoire avait pour mission de former des enseignants du secondaire, la massification a entraîné le développement de sorties professionnelles différentes, non négligeables³⁰, parfois fort éloignées des carrières auxquelles donnait accès traditionnellement ce diplôme, comme des emplois dans le secteur bancaire. Cette diversification des sorties a constitué le déclencheur d'une interrogation sur les acquis des étudiants : correspondent-ils à ce que l'on attend d'eux dans les milieux professionnels ? Les étudiants sont-ils capables d'identifier leurs acquis, de manière à les mettre en avant dans leur démarche de recherche d'emploi ? ...

4.2 LES EFFETS DOMMAGEABLES

Les principaux arguments au nom desquels une amélioration de la qualité des pratiques d'évaluation des acquis pourrait être réclamée sont repris ci-dessous de manière synthétique, sur la base des effets néfastes des pratiques actuelles et de leurs lacunes.

Pour l'État

- L'importante diversité des pratiques d'évaluation des acquis et surtout leur opacité placent l'État dans une situation délicate de contradiction de plus en plus grande entre des prétentions officielles concernant la **valeur nationale des diplômes** et leur hétérogénéité de fait. Bien sûr, le caractère national du diplôme ne contraint pas à une uniformité de contenu. Prétendre que les diplômes ont une valeur nationale ne requiert pas nécessairement que l'évaluation porte sur les mêmes acquis, ce qui serait d'ailleurs contraire à l'idée même de formation universitaire. Le caractère national des diplômes a davantage trait à la garantie accordée par l'État que les procédures locales répondent à des normes de qualité reconnues. Or, on ne voit pas bien comment l'État pourrait apporter sa caution à des procédures opaques, qui n'explicitent pas les objets sur lesquels elles portent.
- La méconnaissance des acquis sur le plan national génère un cruel manque de données pour alimenter tout débat à propos de **l'adéquation entre les compétences de sortie des étudiants et les compétences attendues** par le monde économique et social, notamment dans le cadre de la nouvelle finalité d'employabilité de l'enseignement tertiaire, décrite ci-dessus. Le pilotage national des formations en est rendu difficile (par exemple, le développement d'une politique volontariste d'accroissement de tel type de diplômés en fonction des besoins de telle région).

³⁰ De l'ordre de 20 à 25% des diplômés.

- Depuis la crise des finances publiques des années 80 et au fur et à mesure que l'enseignement universitaire de masse accaparait une partie non négligeable des budgets alloués à l'éducation, il s'est développé, sous différentes formes, une tendance à exiger de ce niveau d'enseignement qu'il rende compte de l'utilisation des fonds publics qui lui sont accordés. Ce souci de **transparence** passe par l'élaboration d'indicateurs clairs de ce que font les universités, c'est-à-dire, entre autres choses, de ce que les étudiants y développent en termes de connaissances et de compétences.
- Alors que le développement économique d'un pays est de plus en plus tributaire de sa capacité à acquérir et à utiliser des connaissances et des compétences de haut niveau, on attend de l'enseignement supérieur qu'il explicite sa contribution à la « **société du savoir** » en identifiant mieux les acquis des étudiants à leur sortie des études.
- La certification locale et contextualisée constitue aussi un **obstacle majeur à l'évaluation du fonctionnement pédagogique** des établissements d'enseignement supérieur (Jarousse et Michaut, 2001). En effet, si l'on cherche à mesurer de manière fiable l'efficacité des modes d'organisation pédagogique, il faudrait disposer d'une mesure de la réussite des étudiants, externe et exogène aux établissements³¹.

Pour le monde économique et social

- La méconnaissance des acquis est tout aussi dommageable pour les futurs employeurs des étudiants : les méthodes de certification normative actuelles **ne les renseignent pas sur les compétences effectivement maîtrisées** par les étudiants à la sortie des études.

Pour les universités

- Dans le cadre de la **création d'un espace européen de l'enseignement supérieur**, les universités les plus à même de relever les défis inhérents à cette création, comme la mobilité et la concurrence, seront sans doute celles qui auront entrepris des efforts d'explicitation de ce que valident leurs diplômes. La lisibilité des formations universitaires, y compris en termes des acquis qu'elles sanctionnent, pourrait s'avérer un facteur déterminant d'une intégration européenne réussie. De manière générale d'ailleurs, un plus grand professionnalisme dans l'évaluation des acquis peut constituer un facteur d'attractivité non négligeable.
- L'explicitation des acquis attendus serait aussi de nature à alimenter un débat, aujourd'hui plus nécessaire qu'hier, sur **l'adéquation entre les formations et les demandes** liées aux nouveaux débouchés. Comme les enseignants-chercheurs nous l'ont rapporté lors de l'étude de cas, le fait que plus d'un quart des étudiants de la filière histoire trouvent en définitive du travail dans le secteur privé a rendu nécessaire une réflexion en profondeur sur la pertinence de leurs acquis en regard de ce nouveau type de trajectoire professionnelle.

Pour les enseignants

- En l'absence de repères explicites et officiels sur le « produit attendu », les enseignants-chercheurs éprouvent des **difficultés à déterminer leur niveau d'exigence**. Ce phénomène s'observe particulièrement en début de carrière, lorsque l'enseignant-chercheur nouvellement recruté prend en charge de nouveaux cours, sans qu'il reçoive (liberté académique oblige...) des indications précises quant à leurs objectifs. Cet effet dommageable est particulièrement important alors que les universités vont connaître un renouvellement important de leur personnel enseignant.

³¹ On pourrait bien sûr rétorquer que l'insertion professionnelle des étudiants constitue une évaluation externe possible de la qualité des acquis des étudiants et, en conséquence, de la qualité des formations. Sans entrer dans le détail, il est possible de montrer que cette évaluation ne serait guère valide. En effet, une mauvaise insertion professionnelle peut être due à un marché du travail local trop étiqué. Une bonne insertion peut résulter de la qualité des étudiants recrutés au départ de la formation. Dans ces deux cas, l'insertion ne constitue pas un indicateur à coup sûr valide de la qualité des formations.

- Une autre difficulté ressentie par les enseignants consiste à devoir **s'adapter à l'hétérogénéité des « nouveaux publics »**, en l'absence d'une vision claire des compétences à atteindre. Autrement dit, comment adapter l'enseignement à une plus grande diversité de compétences d'entrée si on ne dispose pas d'un référentiel de compétences attendues à la sortie ? Une évaluation diagnostique fine à l'entrée ne peut que reposer sur une définition préalable des compétences à développer.
- Surtout au premier cycle, on assiste alors au **développement d'une évaluation normative, pour une part étrangère à la maîtrise de connaissances ou de compétences** dont fait preuve l'étudiant (cf. 3.3).

Pour les étudiants...

- Un déficit d'**équité** : un étudiant peut ne pas obtenir tel diplôme dans tel établissement alors qu'il l'aurait peut-être obtenu sans peine ailleurs.
- Parce qu'ils manquent de repères quant aux objectifs à atteindre, les étudiants éprouvent des difficultés à **orienter leurs pratiques d'étude** et à déterminer ce qu'ils doivent mettre en œuvre pour réussir.
- Le manque de clarification des critères peut aussi être tenu pour un des facteurs qui entrave la **démocratisation** de l'accès et de la réussite au supérieur. En effet, les étudiants qui disposent, grâce à leur environnement familial notamment, des clés de décodage des attentes implicites du supérieur, ont davantage de chance passer le cap du premier cycle.
- De plus, la **motivation** des étudiants et leur **engagement dans un apprentissage en profondeur** semblent être favorisés par la transparence de l'évaluation. L'opacité des pratiques d'évaluation les amène à concevoir les examens comme un « jeu du chat et de la souris ». Ils consacrent alors l'essentiel de leur énergie à en comprendre les règles, plus qu'à réaliser des apprentissages de qualité. Les examens étant, à leurs yeux, une suite de pièges à déjouer, ils passent une bonne partie de leur temps à chercher ce que les différents enseignants attendent d'eux.
- Les étudiants ne connaissent pas les domaines dans lesquels la formation leur a assuré des acquis et ne peuvent donc pas les mettre en avant dans leur **relation avec leurs futurs employeurs**.

5 COMMENT AMÉLIORER LES PRATIQUES D'ÉVALUATION DES ACQUIS ?

L'objectif d'une mission d'expertise est de dresser un état des lieux et de poser un diagnostic. Formuler des recommandations dépasse sans doute la stricte mission de l'expert. Il nous a cependant paru utile, dans la dernière partie de ce rapport, de tracer les grandes lignes des actions qui pourraient être entreprises pour améliorer l'évaluation des acquis des étudiants. Ces propositions n'ont d'autre prétention que d'ouvrir le débat sur les implications éventuelles du présent rapport.

5.1 LE DÉVELOPPEMENT D'ÉPREUVES STANDARDISÉES ET INDICATIVES D'ÉVALUATION DES ACQUIS

Une première manière de lutter contre l'hétérogénéité des pratiques actuelles d'évaluation et de combler l'absence de données sur cette question consisterait à développer et à expérimenter des épreuves qui ont fait l'objet d'une forme, plus ou moins poussée, de standardisation. Selon une première modalité, il pourrait s'agir d'**encourager la collaboration** entre les différents responsables d'une même filière, de manière à élaborer conjointement des épreuves communes ou des parties d'épreuves communes, comme dans le cas des épreuves *Message* rapporté au point 3.2. Sans aller jusque-là, la collaboration peut déjà produire des effets bénéfiques, si elle cherche à comparer les pratiques d'évaluation de manière à identifier les points communs et les différences entre les standards académiques des différentes équipes, comme dans l'expérience originale de création, sur le mode d'une nouvelle pratique de management très en vogue (le « benchmarking³² »), d'un « club de benchmarking » pour les pratiques d'évaluation de la filière d'histoire de différentes universités anglaises (Woolf & Cooper., 1999). L'objectif de ce « club », dans le respect de l'autonomie de ses membres, est d'identifier les pratiques d'évaluation des différents établissements, d'explorer les arguments qui les justifient, de comparer les résultats d'étudiants inscrits dans des modules identiques, à un même niveau d'études, de manière à identifier les standards académiques qui prédominent dans chaque institution.

Selon un deuxième scénario, sans encore chercher à être représentatif de ce qui se passe dans chaque établissement, des recherches pourraient être menées sur des échantillons restreints, dans le but d'identifier les types d'acquis que les études universitaires assurent, sur le modèle des recherches américaines consacrées aux effets de l'enseignement supérieur. En particulier, on pourrait suggérer que les **recherches en éducation** s'intéressent plus régulièrement au supérieur, en consacrant notamment des études aux effets de ce niveau du système scolaire, dont on connaît mal les retombées en termes de maîtrise de connaissances et de compétences.

À plus grande échelle, il pourrait enfin s'agir d'**épreuves indicatives nationales**, construites par un organisme indépendant sur la base, d'une part, des maquettes nationales et, d'autre part, de relevés des points communs qui apparaissent entre les objectifs annoncés par les différents établissements. On pourrait ici s'inspirer des méthodologies mises en œuvre dans les programmes internationaux d'évaluation des compétences des élèves de différents pays (par exemple, le récent programme PISA) ainsi que de celles qui ont été développées pour les autres niveaux scolaires, notamment en France par la Direction de la programmation et du développement. À la différence des opérations menées aux autres niveaux scolaires, il ne s'agirait pas d'évaluations de masse visant à mesurer, pour tel diplôme, les acquis de tous les étudiants, étant donné l'autonomie pédagogique des établissements et la diversité des programmes qu'elle entraîne. Il ne s'agirait pas non plus de l'évaluation des établissements, qui relève du CNE. Ces épreuves, menées sur des échantillons représentatifs d'étudiants, fourniraient un portrait indicatif des connaissances et des compétences acquises par les étudiants de telle filière, à tel palier de leurs études. Elles seraient, bien entendu, distinctes des procédures de certification, qui

³² Le « benchmarking » consiste, pour une entreprise, à bien analyser ses pratiques (processus et résultats), à les comparer, en interne, entre ses différents établissements et, en externe, avec les meilleurs de ses concurrents, dans son pays ou à l'étranger, et à agir pour les évaluer, voire les dépasser.

doivent rester aux mains des équipes locales, compte tenu de l'autonomie pédagogique dont elles disposent. Il devrait être clairement établi au départ que la fonction principale de ce portrait est que ces équipes locales s'en saisissent pour, en interne, nourrir leur réflexion, d'une part, sur leurs propres objectifs de formation en regard de cet inventaire et, d'autre part, sur les modalités à mettre en œuvre pour améliorer la qualité de leurs enseignements. L'effet visé, respectueux de l'autonomie des acteurs, réside dans ce que Joutard et Thélot (1999) appellent « l'effet miroir » : « *C'est au vu des éléments que l'évaluation a mis en évidence, dont, grâce à elle, l'acteur prend conscience plus nettement, l'évaluation jouant le rôle de miroir précisément, qu'il modifiera, si nécessaire, ses pratiques pour les rendre plus efficaces* » (Joutard et Thélot, 1999, p. 132).

Malgré ces garanties, il faudrait méconnaître sérieusement le monde académique pour ne pas prédire que cette dernière proposition soulèvera une véritable levée de boucliers. Le principe même de ce type d'épreuves est en effet difficilement acceptable a priori par les universitaires, « *peu habitués à un quelconque contrôle pédagogique* » (Duru-Bellat et al, 2000, p.143). Les arguments hostiles ne manqueront pas : ils porteront, pour l'essentiel, sur le fait que le développement de ce type d'épreuves mettrait à mal l'idée même d'université. En effet, la spécificité et la qualité de l'enseignement universitaire sont souvent présentées comme tributaires de la liberté pédagogique, dont jouissent les équipes locales, de tirer parti des spécialités en recherche de leurs enseignants et de la marge d'autonomie qui leur permet d'adapter la formation à son environnement (public étudiant, demandes locales de compétences, ...).

Une première manière de répondre à ces arguments est qu'il vaut peut-être mieux que le monde universitaire participe à l'organisation sérieuse et rigoureuse de ce type d'épreuves, aux conditions qu'il aura lui-même fixées, plutôt que de se voir imposer des évaluations sauvages et incontrôlées qui ne manqueront pas de naître dans un « marché » européen de l'enseignement supérieur. Par ailleurs, en fonction des effets dommageables décrits au point précédent et de la pression sociale qui s'exerce sur l'ensemble des systèmes éducatifs pour qu'ils rendent compte de l'utilisation des fonds publics, on ne voit pas par quel miracle le supérieur serait le seul niveau à échapper à ces bilans d'acquis. Enfin et surtout, il s'agirait de montrer que le développement d'épreuves standardisées, telles que nous les avons décrites, n'est pas incompatible avec la liberté académique et qu'il peut, tout au contraire, constituer un instrument d'affirmation de son indépendance et une manière de mettre en évidence ses spécificités (par exemple, en se proposant de mettre l'accent sur des compétences peu développées ailleurs). Dans le cadre de l'espace européen de l'enseignement supérieur, ces épreuves indicatives seraient d'ailleurs de nature à mieux positionner les universités qui pourraient ainsi rendre plus lisibles leurs formations et expliciter davantage les connaissances et les compétences que les étudiants peuvent y acquérir. Comme le signalait un directeur d'UFR lors d'un entretien, « *l'université ne sait pas se vendre* » : ainsi, on ne sait pas précisément pourquoi les étudiants universitaires continuent à être appréciés sur le marché du travail, malgré l'augmentation des diplômés. Sans doute est-ce parce qu'ils ont acquis à l'université des compétences et des attitudes intellectuelles de haut niveau, mais sans que ces dernières soient identifiées et validées en tant que telles. Des bilans d'acquis indicatifs pourraient aider à y voir plus clair.

Si l'on veut conserver la moindre chance que de pareilles épreuves voient le jour, il conviendrait donc de leur fixer des balises fermes, qui tiennent compte des fortes résistances que ce projet ne manquera pas de rencontrer, mais aussi des risques et des effets pervers possibles de ce type d'opération. Voici une première liste de ces balises.

- Le développement de ces épreuves doit, à chaque stade, être réalisé dans le respect de l'autonomie pédagogique des établissements (de la définition des champs de connaissances et de compétences à mesurer jusqu'aux modes de correction des épreuves).
- Ces épreuves ne doivent pas chercher à tester des connaissances ou des compétences trop spécifiques, puisque les programmes sont assez hétérogènes. Elles devraient davantage porter sur des compétences générales et de haut niveau, que les différents établissements cherchent à atteindre, au-delà de la diversité des parcours de formation.

- Les épreuves devraient être d'emblée conçues, non comme des instruments de contrôle³³ et encore moins de mise en concurrence des établissements entre eux, mais dans une perspective de régulation et d'amélioration des pratiques pédagogiques, comme nous l'avons décrit ci-dessus. Elles constitueraient des photographies, des bilans indicatifs de connaissances et de compétences dont les équipes pédagogiques peuvent s'emparer pour alimenter leur réflexion sur l'amélioration de leurs formations. Concrètement, cette exigence implique que les équipes locales soient les premiers destinataires et les premiers bénéficiaires des résultats de ces bilans et qu'elles gardent la mainmise sur l'utilisation et la diffusion de ces résultats.
- Ces épreuves, à visée informative, ne se substituent pas à la certification locale. Elles sont d'ailleurs réalisées ponctuellement, sur des échantillons représentatifs d'étudiants seulement. Elles pourraient, par exemple, être organisées en parallèle à l'évaluation des filières que mène le CNE.

5.2 LA PROFESSIONNALISATION DE L'ÉVALUATION LOCALE DES ACQUIS

L'autonomie institutionnelle et l'attachement des universitaires à la liberté académique font que les actions évoquées au point précédent sont sans doute appelées à rester ponctuelles et modestes, si tant est qu'elles voient le jour. Une deuxième piste, plus réaliste, d'amélioration de l'évaluation des acquis, réside dans le développement, au sein même de chaque établissement et dans le respect de son autonomie, d'un souci accru de professionnaliser les actes d'évaluation. Cet objectif général peut être atteint par de multiples voies. Citons les principales.

5.2.1 Le développement d'une culture institutionnelle de la qualité autour de l'évaluation

En interne, la qualité des pratiques d'évaluation des acquis devrait constituer une des dimensions de la démarche qualité globale de l'établissement. La question de l'évaluation devrait donc figurer en meilleure place au sein de la liste des aspects à envisager quand un établissement souhaite faire le point sur son fonctionnement pédagogique, que ce soit spontanément ou à l'occasion de la contractualisation, de la définition de son projet d'établissement ou de la rédaction du rapport d'auto-évaluation préparatoire aux visites de site du CNE. De ce point de vue, l'excellent guide d'auto-évaluation produit par l'AMU³⁴ constitue un outil appréciable. En particulier, la fiche 5 D fournit une liste détaillée de questions d'autoévaluation à propos des pratiques d'évaluation des acquis. L'Agence anglaise de l'assurance-qualité pour l'enseignement supérieur a également élaboré un « code de conduite » sur ce même thème³⁵.

Des expériences étrangères en la matière, on peut tirer un certain nombre de principes qui gouverneraient un pareil développement d'une démarche qualité interne autour des pratiques d'évaluation des acquis (Aper & Hinkle, 1991) :

- veiller à l'implication la plus large possible de l'ensemble des enseignants ;
- faire expliciter au préalable les objectifs éducatifs généraux de l'institution et les visées plus particulières des diverses formations en termes d'acquis des étudiants ;
- établir et mettre en œuvre des procédures qui permettent de savoir si ces objectifs sont atteints ;
- utiliser les données recueillies pour réguler les pratiques d'évaluation des acquis ;

³³ En particulier, il faudrait veiller au risque que ces bilans d'acquis soient utilisés comme critères d'appréciation de l'efficacité des établissements ou des équipes pédagogiques. En effet, les acquis cognitifs des étudiants constituent des critères d'évaluation trop étroits par rapport à l'ensemble des missions que remplit actuellement l'université, comme l'orientation progressive (Duru-Bellat et al., 2000).

³⁴ Réalisé dans le cadre des travaux du Comité scolarité et vie étudiante de cette agence, le guide est disponible à l'adresse suivante : <http://www.cpu.fr/Dossier/GuideScol/>

³⁵ Disponible sur le site de l'Agence (<http://www.qaa.ac.uk/public/COP/codesofpractice.htm>)

- insérer des indicateurs d'acquis dans les évaluations des enseignements par les étudiants.

On ne saurait trop insister sur ce dernier principe. En effet, alors que se développe actuellement l'évaluation des enseignements par les étudiants, il est à regretter que les indicateurs utilisés soient trop souvent centrés sur des aspects assez triviaux et de surface des processus d'enseignement, laissant dans l'ombre leurs résultats. Un cours clair et donné avec enthousiasme ne constitue pas encore une garantie - même si ces qualités sont bien sûr importantes- que des acquis pertinents ont été installés. Il s'agirait donc de développer des questionnaires d'évaluation qui recueilleraient aussi des informations sur les effets de l'enseignement³⁶ (Dejean, 2002 ; Pratt, 1997).

5.2.2 La formation des enseignants-chercheurs à l'évaluation

La deuxième piste de développement d'une culture institutionnelle de la qualité autour de l'évaluation des acquis a trait à la formation pédagogique des enseignants-chercheurs. Une formation de base aux mécanismes de l'évaluation pourrait prendre place en formation initiale, par exemple sous la forme d'un module à insérer dans les formations prodiguées dans les CIES ou dans les modules d'accueil pédagogique des enseignants-chercheurs nouvellement recrutés que développent les universités. Ultérieurement, une formation continue pourrait être destinée aux enseignants en fonction, sur le mode d'ateliers de réflexion sur leurs propres pratiques d'évaluation des acquis³⁷.

Les objectifs de ces formations pourraient être les suivants :

- sensibiliser les enseignants aux problèmes docimologiques les plus fréquemment rencontrés dans le supérieur (cf. 3.4) et les accompagner dans l'élaboration de garde-fous à mettre en place pour professionnaliser davantage l'évaluation des étudiants ;
- fournir des outils de décodage du fonctionnement du jugement évaluatif ;
- discuter du statut et des limites des outils d'évaluation, comme les échelles de notation, de manière à adopter une attitude plus distante et plus relativiste par rapport à ces outils ;
- porter un regard critique sur ses propres pratiques d'évaluation selon les questions suivantes :
 - l'évaluation se réalise-t-elle en rapport à des objectifs explicites de formation ?
 - les tâches proposées dans les épreuves d'évaluation sont-elles bien de nature à mesurer l'atteinte de ces objectifs ?
 - l'évaluation est-elle congruente par rapport à l'enseignement, quant à sa forme et quant à son contenu ?
 - peut-on assurer un degré satisfaisant de fidélité et de fiabilité aux procédures d'évaluation, en particulier à la correction ?
 - les étudiants ont-ils été informés de la forme, du contenu et des critères d'évaluation ? Quels dispositifs pédagogiques sont développés pour que les étudiants s'approprient ces informations ?
 - les procédures mises en place respectent-elles les règlements externes et internes ?

³⁶ Notamment pour assurer une meilleure validité aux évaluations de l'enseignement réalisées par les étudiants. On sait en effet que cette évaluation, comme tant d'autres, n'est pas à l'abri de l'effet de halo : les étudiants sont parfois influencés, dans leur jugement sur la qualité de l'enseignement, par le charisme de l'enseignant (Shevlin et al., 2000).

³⁷ À titre indicatif, on consultera les programmes de formation de l'Institut de pédagogie et du multimédia de l'Université de Louvain (<http://www.ipm.ucl.ac.be/Formation/Agenda.html>) ainsi que le site du Service de pédagogie universitaire de l'Université de Namur et notamment le numéro 47 de la revue pédagogique interne RÉSEAU consacré à la présentation d'un guide pratique pour élaborer et corriger un examen (<http://www.det.fundp.ac.be/spu/reseau.html>)

5.2.3 La clarification du « contrat pédagogique »

Une troisième piste consiste à encourager les établissements et les équipes d'enseignants à expliciter, autant que faire se peut, les règles qui président à l'évaluation des acquis. L'opacité de ces règles, on l'a vu, transforme en effet l'évaluation en « jeu du chat et de la souris », particulièrement préjudiciable aux étudiants défavorisés qui ne disposent pas des outils de décodage des « allants de soi » universitaires. Un double travail est ici nécessaire. Au niveau de l'établissement d'abord, il s'agit de définir et de porter à la connaissance des étudiants les règles générales du contrôle des connaissances et les règles spécifiques à chaque filière et année d'études. C'est ainsi que de nombreuses universités publient désormais des « Chartes des examens »³⁸ et éditent des recueils de modalités particulières.

Mais il appartient aussi à chaque enseignant d'explicitier ses exigences spécifiques, notamment en communiquant aux étudiants, dès le début de l'année, une série d'informations utiles à la compréhension de la logique du cours : les objectifs, la structure du cours, les connaissances préalables requises, la charge de travail de l'étudiant, la quantité et le type d'effort nécessaire à chaque étape, les règles et procédures d'examen, le type de questions, les critères de correction, leur importance relative, etc. Les universités québécoises ont formalisé cette démarche dans ce qu'elles appellent le « plan de cours », qui spécifie le contrat pédagogique et qui constitue une exigence administrative que chaque titulaire doit rencontrer (Langevin & Bruneau, 2000).

Enfin, tout pédagogue sait bien que dire les choses ne suffit pas toujours à ce qu'elles soient comprises. De plus, l'explicitation et la communication des critères d'évaluation n'ont, de l'avis même des étudiants, qu'un impact limité si ces opérations ne s'intègrent pas dans une démarche pédagogique globale, incluant une explication détaillée de ces critères à partir de cas réels, une possibilité d'en discuter avec les enseignants, une clarification de ce que les termes utilisés dans la grille de critères signifient (O'Donnovan et al., 2000). Il s'agit alors de proposer aux étudiants des situations didactiques actives qui leur permettent de s'appropriier ces informations à propos des objectifs et des exigences, comme la simulation d'examens, la simulation de corrections d'examens ou encore la comparaison d'examens antérieurs, réussis et échoués. Ces activités pédagogiques, réalisées ponctuellement mais tout au long du cours, favorisent la construction d'une représentation partagée des critères d'évaluation. Elles permettent aussi aux étudiants d'observer les critères effectifs, qui peuvent parfois (cf. 3.4) se révéler assez différents des critères annoncés.

Ce travail d'explicitation préalable des critères est aussi de nature à augmenter la fidélité des évaluations dans le cas de correcteurs multiples, même si, en elle-même, cette explicitation n'assure pas automatiquement un haut degré d'accord entre correcteurs. En effet, des études fines sur les processus évaluatifs à l'université montrent que la diversité de statuts et d'expériences des évaluateurs (du professeur titulaire à temps plein jusqu'au chargé de cours horaire, issu du secondaire, à temps réduit) nécessite une socialisation plus poussée, par la discussion entre pairs, de manière à favoriser la construction progressive d'un même sens à attribuer à ces critères (Ecclestone, 2001).

Bien sûr, on doit aussi veiller à ce que l'évaluation reste gérable. Il ne s'agit donc pas de multiplier à l'infini les critères d'évaluation, de les préciser longuement ou encore de produire des guides minutieux sur la manière de les interpréter. On veillera à s'en tenir à un nombre raisonnable de critères clairement explicités et à propos desquels sont organisés des échanges, entre collègues d'abord au sein d'équipes pédagogiques, entre l'enseignant et ses étudiants ensuite, par exemple sur la base d'une comparaison entre des « bonnes » et des « mauvaises » copies d'années antérieures.

³⁸ Le guide de l'AMU évoqué ci-dessus fournit, dans la fiche consacrée au contrôle des connaissances, une liste de sites sur lesquels on peut trouver des exemples de chartes des examens (<http://www.cpu.fr/Dossier/GuideScol/>).

5.2.4 Le développement de pratiques innovantes d'évaluation des acquis

S'il est vrai que « le curriculum réel, c'est l'évaluation », c'est la nature même des épreuves qui serait à modifier si l'on souhaite améliorer la qualité des acquis, dans le sens des défis évoqués tout au long du présent rapport. À plusieurs reprises, on a en effet souligné que les pratiques actuelles d'évaluation ne semblent pas toujours propices à développer chez les étudiants les compétences de haut niveau que l'université a de tout temps cherché à promouvoir et que la société du savoir réclame d'elle, aujourd'hui plus qu'hier. On a aussi montré que les compétences attendues par les employeurs et les compétences qui sont considérées par les étudiants comme les principaux acquis de leur passage à l'université ne sont pas toujours celles qui sont en jeu dans les examens universitaires traditionnels (2.2 et 3.5.2). Ces compétences complexes liées au savoir-faire cognitif et aux attitudes, qui sont déjà développées par les études universitaires (2.2), le seraient encore davantage si elles constituaient le cœur des pratiques d'évaluation des acquis, puisque les étudiants sont « pilotés » par l'évaluation (3.5). On comprend alors pourquoi les innovations en matière d'évaluation constituent une part importante de la pédagogie universitaire, au point qu'une revue leur soit consacrée³⁹. Les innovations dans les pratiques d'évaluation prennent les directions suivantes.

Une première évolution majeure a trait au passage d'une **évaluation sanction**, « en bout de course », à une **évaluation formative**, conçue comme partie intégrante de l'apprentissage (cf. 3.2.4). Selon Dochy et al. (1999), l'ère du « testing » fait désormais place, dans le supérieur, à celui de l'évaluation formative.

Tableau 5
Deux conceptions de l'évaluation dans le supérieur
(d'après Dochy et al., 1999)

Ère du testing	Ère de l'évaluation formative
Séparation entre l'enseignement et l'évaluation	Intégration de l'évaluation à l'apprentissage
Évaluation subie par les étudiants	Participation active de l'étudiant
Mesure de connaissances décontextualisées et sans rapport avec l'expérience des étudiants	Mesure de compétences sur des tâches authentiques et complexes
Dont le produit se résume à une note	Dont le produit premier est le processus

Un des défis majeurs réside dans la mise en place de nouvelles formes d'évaluation formative adaptées aux grands groupes, comme le recours à l'autoévaluation critériée : les étudiants sont entraînés à utiliser des grilles de critères qu'ils appliquent à leurs propres productions ou à celles de leurs pairs.

La deuxième évolution concerne le développement d'une **évaluation critériée davantage que normative**. La préoccupation première de l'enseignant devrait être de chercher à savoir ce que chaque étudiant a appris par rapport à ce qu'il devait apprendre et non pas s'il a appris plus ou moins que ses pairs. Cela suppose que l'évaluation se réalise par rapport à des objectifs d'enseignement clairement explicités au préalable et à partir desquels les critères de réussite sont définis et annoncés. L'évaluation sanctionne l'atteinte des objectifs de la formation qu'elle clôt, ce qui présuppose que ces derniers aient été définis clairement au préalable.

³⁹ La revue « Assessment & Evaluation in Higher Education » publie en effet de très nombreux comptes rendus de pratiques innovantes d'évaluation. L'index des dernières années est consultable sur le site suivant : <http://www.tandf.co.uk/journals/archive/c-archive/ae-con.html>

La troisième évolution touche à la nécessité de **diversifier les méthodes d'évaluation** et notamment d'y intégrer des pratiques nouvelles qui soient **compatibles avec l'apprentissage en profondeur et avec le développement de compétences de haut niveau**, comme l'autonomie, le travail de groupe, l'esprit critique (Brown et al., 1994 ; Dochy et al., 1999 ; Fallows & Chandramohan, 2001 ; Montgomery, 1995 ; Wisker, 1997).

- C'est ainsi que l'étudiant est impliqué dans les pratiques d'évaluation (**autoévaluation** ou **évaluation par les pairs**⁴⁰), notamment pour favoriser le développement de son autonomie.
- L'évaluation ne se réduit plus à une mesure ponctuelle, mais vise au contraire à s'élaborer sur le long terme et de manière personnelle, par exemple via le **portfolio** (Goupil et al., 2000).
- **L'évaluation par compétences** est aussi présentée comme une manière de dépasser les lacunes des pratiques actuelles d'évaluation des acquis, même si, d'une part, il faut reconnaître qu'il ne s'agit pas de la panacée au point de vue docimologique et si, d'autre part, elle semble appropriée à certains types de programmes plus qu'à d'autres, en particulier les formations à orientation professionnelle (Edwards & Knight, 1995).
- Le recours aux méthodes actives exige aussi une révision en profondeur des procédures d'évaluation et l'introduction de **méthodes d'évaluation congruentes avec les principes sur lesquels se fondent ces pédagogies actives**. Par exemple, le talon d'Achille de l'apprentissage par problèmes, introduit dans de nombreuses facultés de médecine et d'ingénieur, a longtemps été la discordance entre ses principes méthodologiques innovants (apprentissage collaboratif, inductif, ...) et les procédures d'évaluation qui restaient traditionnelles. Les étudiants estiment que cette méthode, qui se fonde pour l'essentiel sur le traitement actif et en groupe de cas concrets authentiques ou semi-authentiques, ne les prépare pas à réussir les examens classiques, qui restent, quant à eux, centrés sur la restitution de connaissances. Des méthodes d'évaluation plus adéquates aux fondements et aux finalités de ces méthodes actives ont donc dû être mises au point (Segers et Dochy, 2001). Par exemple, l'évaluation se réalise à partir de situations authentiques, ou du moins proches de la complexité des situations réelles et fait intervenir le travail de groupe.
- Enfin, les procédures d'évaluation devraient davantage soigner **les modes de communication des résultats**, de manière notamment à mieux renseigner les étudiants et leurs futurs employeurs sur les acquis effectivement maîtrisés en fin de formation. À titre d'exemple, Potterton & Parsons (1995) ont développé un projet original visant à communiquer les résultats de l'évaluation en mentionnant les compétences acquises et leur niveau d'acquisition pour chaque étudiant.

Il conviendrait enfin que les expériences innovantes en matière d'évaluation soient portées à la connaissance du plus grand nombre, via par exemple, des inventaires de « bonnes pratiques » qui pourraient être réalisés dans le cadre de l'AMU.

5.3 UNE PLUS GRANDE PRISE EN COMPTE DES PRATIQUES D'ÉVALUATION DES ACQUIS DANS LES PROCÉDURES DE RÉGULATION

Cette dernière série de mesures a trait au rôle de l'État et aux moyens par lesquels les pouvoirs publics pourraient inciter les universités à améliorer l'évaluation des acquis. Si l'on s'accorde à dire que le rôle des autorités publiques, compte tenu de l'autonomie des établissements, ne consiste pas à vérifier que tous les étudiants d'une même filière reçoivent le même enseignement et développent

⁴⁰ Ce qui ne manque d'ailleurs pas de poser de nouvelles questions docimologiques de fidélité et de validité (pour une synthèse, Dochy et al. 1999). Ainsi, la fidélité d'évaluations de compétences à l'oral, réalisées par des pairs, semble plus faible que celles effectuées par des enseignants. Pour assurer le même niveau de fidélité dans le temps obtenu par un seul enseignant, même si ce niveau reste modéré, il faut recueillir l'évaluation d'au moins 2 à 4 étudiants et en faire la moyenne (Magin & Helmore, 2001).

donc les mêmes acquis⁴¹, il appartient quand même à ces autorités de vérifier que l'évaluation des acquis réponde aux standards et aux critères de qualité attendus et qu'elle se déroule, dans chaque établissement, de manière valide, fidèle, crédible et fiable. Autrement dit, le rôle de l'État n'est pas tant, en ce qui concerne l'enseignement supérieur, d'assurer une qualité égale des acquis que de donner sa caution sur la qualité des processus mis en œuvre pour valider localement ces acquis. Les efforts devaient donc davantage porter sur l'explicitation et l'analyse de la qualité des pratiques locales que sur leur uniformisation.

Dans cette perspective, un certain nombre d'organismes chargés de réguler les formations universitaires seraient à même de donner un message fort selon lequel la qualité de l'évaluation des acquis des étudiants constitue un aspect important du fonctionnement pédagogique des établissements universitaires.

- Ainsi, **l'évaluation des établissements et des filières**, telle qu'elle se pratique par le CNE, pourrait concerner davantage les pratiques d'évaluation des acquis, par exemple en demandant que les rapports d'autoévaluation s'interrogent sur le type d'évaluation privilégié et sur sa cohérence par rapport aux objectifs poursuivis.
- Il en va de même pour les **procédures d'habilitation** à délivrer des diplômes nationaux, mises en place par le Ministère de l'Éducation nationale. La description des éléments liés à l'évaluation, de l'explicitation des objectifs des formations jusqu'aux procédures du contrôle des connaissances, pourrait être davantage prise en compte dans l'appréciation de chaque dossier.
- Cet aspect du fonctionnement pédagogique pourrait aussi être mis en avant dans les procédures de **contractualisation**. En particulier, on pourrait attendre du projet d'établissement qu'il précise les actions menées pour améliorer la qualité de l'évaluation des acquis. Le Ministère pourrait ainsi établir un recueil de pratiques innovantes en la matière et contribuer à sa diffusion.
- Enfin, la nécessité européenne de concevoir une annexe descriptive aux diplômes, appelée « **supplément au diplôme** », pourrait être l'occasion d'une réflexion sur le contenu à exiger d'un tel document en termes d'explicitation des connaissances et des compétences acquises. Un décret du 8 avril 2002 portant sur les implications de la construction de l'Espace européen de l'enseignement supérieur stipule en effet que ce supplément au diplôme devrait permettre d'assurer, dans le cadre de la mobilité, une meilleure « *lisibilité des connaissances et aptitudes acquises* ».

⁴¹ Ce qui n'exclut pas des coups de sonde indicatifs et formatifs, tels qu'ils ont été proposés au point 5.1.

6 RÉFÉRENCES

- Aper J. & Hinkle D. (1991). State policies for assessing student outcomes. *Journal of Higher education*, 62, 5, 539-555.
- Ashcroft K. & Palacio D. (1996). *Researching into assessment and evaluation in colleges and universities*. London : Kogan Page.
- Astin A. (1991). *Assessment for excellence. The philosophy and practice of assessment and evaluation in higher education*. New York : MacMillan Publishing Company.
- Beattie J.-F. (1995). Évaluation dans l'enseignement supérieur. *Gestion de l'enseignement supérieur*, 7 (3), 307-324.
- Bertrand D., Foucher R., Jacob R., Fabi B. & Beaulieu P. (1994). *Le travail professoral remesuré. Unité et diversité*. Sainte-Foy : Presses de l'Université du Québec.
- Biggs J. (1999). *Teaching for quality at University*. London : Open University Press.
- Bireaud A. (1996). En France, une politique de formation pédagogique pour les enseignants du supérieur timide, hésitante et controversée. In J.Donnay & M.Romainville (eds). *Enseigner à l'université : un métier qui s'apprend ?* Bruxelles : De Boeck, 113-122.
- Blais J.-G., Laurier M., Van der Maren J.-M., Gervais C., Lévesque M., Pelletier G. (1997). *L'évaluation des apprentissages à l'Université de Montréal et dans ses écoles affiliées*. Montréal : Groupe de Recherche Interdisciplinaire en Pédagogie Universitaire.
- Boud D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15 (1), 101-111.
- Boulet A., Savoie-Zajc L. & Chevrier J. (1996). *Les stratégies d'apprentissage à l'université*. Sainte-Foy : Presses de l'Université du Québec.
- Bourdieu P. & Passeron J.-C. (1964). *Les héritiers. Les étudiants et la culture*. Paris : Les éditions de minuit.
- Bridges P., Bourdillon B., Collymore D., Cooper A., Fox W., Haines C., Turner D., Woolf H. & Yorke M. (1999). Discipline-related marking behavior using percentages. *Assesment and evaluation in Higher education*, 24 (3), 71-80.
- Brown S. & Glasner A. (eds) (1999). *Assessment matters in higher education*. London : Open University Press.
- Comité national d'évaluation (1995). *Évolution des universités, dynamique de l'évaluation*. Paris : La Documentation française.
- Comité national d'évaluation (1997). *Les missions de l'enseignement supérieur : principes et réalité*. Paris : La Documentation française.
- Comité national d'évaluation (1998a). *La formation des pharmaciens en France – Volume 1 : les études*. Paris : La Documentation française.
- Comité national d'évaluation (1998b). *La formation des pharmaciens en France – Volume 2 : les 24 UFR de pharmacie*. Paris : La Documentation française.
- Comité national d'évaluation (1999). *Enseignement supérieur : autonomie, comparaison, harmonisation. Rapport 1995-1999*. Paris : La Documentation française.
- Comité national d'évaluation (2001). *Rapport d'évaluation : l'université de Limoges*. Paris : La Documentation française.
- Comité national d'évaluation (2002a). *Les formations supérieures en mathématiques orientées vers les applications*. Paris : CNE.
- Comité national d'évaluation (2002b). *Rapport d'évaluation : l'université de Savoie*. Paris : La Documentation française.
- Commission Attali (1998). *Pour un modèle d'enseignement supérieur*. Paris : Stock.
- Commission Boyer (1998). *Reinventing undergraduate education*. New York : Stony Brook.

- Confédération des conférences des recteurs de l'UE (1999). *Évolution des structures d'éducation dans l'enseignement supérieur en Europe*. Bruxelles : Confédération des conférences des recteurs de l'UE.
- Conférence des Recteurs Européens (1997). *Changer l'enseignement supérieur en Europe, un programme pour le XXI^e siècle. Cahiers sur l'Enseignement Supérieur, n° 111*.
- Conférence mondiale de l'UNESCO sur l'enseignement supérieur (1998). *Déclaration mondiale sur l'enseignement supérieur pour le XXI^e siècle : vision et action*. Paris : Unesco.
- Conseil de l'Europe - comité de l'enseignement supérieur et de la recherche (1998). *Projet de recommandation des Ministres aux États membres sur l'accès à l'enseignement supérieur*. Strasbourg : Conseil de l'Europe.
- Conseil supérieur de l'Éducation du Québec (1995). *Réactualiser la mission universitaire*. Québec : Conseil supérieur de l'Éducation du Québec.
- Coulon A. (1997). *Le métier d'étudiant*. Paris : PUF.
- Crahay M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck.
- Dalziel J. (1998). Using marks to assess student performance : some problems and alternatives. *Assessment & evaluation in higher education*, 23, 4, 351-366.
- De Landsheere G. (1980). *Évaluation continue et examens. Précis de docimologie*. Bruxelles : Éditions Labor.
- De Vita G. (2002). Cultural equivalence in the assessment of home and international business management students. *Studies in Higher Education*, 27 (2), 221-231.
- Dearing Committee (1997). *Higher education in a learning society*. London : National Committee of Inquiry into Higher Education.
- Defays J.-M., Maréchal M. & Melon S. (2000) (eds) *La maîtrise du français : du niveau secondaire au niveau supérieur*. Bruxelles : De Boeck.
- Dejean J. (2002). *L'évaluation de l'enseignement dans les universités françaises. Rapport établi à la demande du Haut Conseil pour l'évaluation de l'école*. Paris : Haut Conseil pour l'évaluation de l'école.
- Dejean K. & Magoga E. (2001). *Maîtrise de la langue et échec en première candidature – Rapport de recherche*. Namur : Facultés Universitaires de Namur.
- Dochy F., Segers M. & Sluijsmans D. (1999). The use of self, peer and co-assessment in higher education : a review. *Studies in Higher Education*, 24 (3), 331-350.
- Donnay J. & Romainville M. (eds) (1996). *Enseigner à l'université : un métier qui s'apprend ?* Bruxelles : De Boeck.
- Drew S. (1998). Students' perceptions of their learning outcomes. *Teaching in Higher Education*, 3 (2), 197-217.
- Duru-Bellat M., Jarousse J.-P., Leroy-Audouin C. et Michaut C. (2000). Écueils et enjeux de l'évaluation de l'enseignement supérieur. *Administration et éducation*, 86 (2), 133-146.
- Ecclestone K. (2001). « I know a 2:1 when I see it » : understanding criteria for degree classifications in franchised university programmes. *Journal of Further and Higher Education*, 25 (3), 301-313.
- Edwards A. & Knight P. (1995). *Assessing competence in higher education*. London : Kogan Page.
- Elton L. (1998). Are UK degree standards going up, down or sideways ? *Studies in Higher Education*, 23 (1), 35-42.
- Fallows S. & Chandramohan B. (2001). Multiple approaches to assessment : reflections on use of tutor, peer and self-assessment. *Teaching in Higher Education*, 6 (2), 229-246.
- Fitoussi J.-P. (2001). *L'enseignement supérieur des sciences économiques en question*. Paris : Fayard.
- Franklyn-Stokes A. & Newstead S. (1995). Undergraduate cheating : who does what and why ? *Studies in Higher Education*, 20 (2), 159-172.
- Gibbs G. & Lucas L. (1997). Coursework assessment, class size and student performance. *Journal of further and higher education*, 21 (2), 183-192.
- Gibbs G. (ed) (1995). *Improving student learning through assessment and evaluation*. Oxford : Oxford Brookes University.

- Girod de l'Ain B. (1997). L'avenir des universités européennes. *Gestion de l'enseignement supérieur*, 9 (1), 93-113.
- Goldschmid M. L. (1992). *La réussite professionnelle des diplômés universitaires*. Actes du Congrès de l'Association Internationale de Pédagogie Universitaire de Laval. Québec : Université Laval, 417-425.
- Goupil G., Lussier-Desrochers D., Berthiaume C., Legault A. & Hénault I. (2000). Le portfolio : un exemple d'application dans un cours de cycle supérieur. *Res Academica*, 18, 191-208.
- Greer L. (2001). Does changing the method of assessment of a module improve the performance of a student ? *Assessment and evaluation in higher education*, 26 (2), 127-138.
- Hart G., Bowden J. & Watters J. (1999). Compétences des diplômés. *Enseignement supérieur en Europe*, XXIV (2), 184-194.
- Haug G. (2001). L'employabilité en Europe, dimension-clé du processus de convergence vers un espace universitaire européen. *Politiques d'éducation et de formation*, 2 (2), 11-26.
- Heywood J. (2000). *Assessment in Higher Education*. London : Jessica Kingsley Publishers.
- Husbands C. (1976). Ideological bias in the marking of examinations. *Research in higher education*, 15, 17-38.
- Hutmacher W. (2001). L'université et les enjeux de la professionnalisation. *Politiques d'éducation et de formation*, 2 (2), 27-48.
- Jarousse J.-P. et Michaut C. (2001). Variété des modes d'organisation des premiers cycles et réussite universitaire. *Revue française de Pédagogie*, 136, 41-51.
- Johnston S. (1997). Examining the examiners : a analysis of examiners' rapports on doctoral thesis. *Studies in Higher Education*, 22 (3), 333-347.
- Joutard P. & Thélot C. (1999). *Réussir l'école*. Paris : Seuil.
- Kangis P. (2001). Presentational dimensions and marks awarded to assignments. *Quality in higher education*, 7 (3), 199-206.
- Kehm B. (2001). Oral examinations at German universities. *Assessment in education*, 8 (1), 25-31.
- Kletz F. & Pallez F. (2001). *L'offre de formation des universités : création de diplômes et stratégies d'établissements*. Paris : École des Mines de Paris, Centre de gestion scientifique.
- Knight P. (2002). Summative assessment in higher education : practices in disarray. *Studies in higher education*, 27 (3), 276-286.
- Lahire B. (1997). *Les manières d'étudier*. Paris : la Documentation française.
- Langevin L. & Bruneau M. (2000). *Enseignement supérieur : vers un nouveau scénario*. Paris : ESF éditeur.
- Lebrun M. & Lega J. (1999). Comment mettre en évidence et développer chez l'apprenant les compétences transversales requises pour le préparer à l'enseignement supérieur. *Res academica*, 17 (1et2), 23-40.
- Lessard C. & Bourdoncle R. (2002). Qu'est-ce qu'une formation professionnelle universitaire ? *Revue française de pédagogie*, 139, 131-153.
- LindBlom-Ylänne S. & Lonka K. (2001). Students'perceptions of assessment practices in a traditional medical curriculum. *Advances in health sciences education*, 6, 121-140.
- Little A. & Wolf A. (1996). *Assessment in transition. Learning, monitoring and selection in international perspective*. Oxford : Elsevier Science Ltd.
- Lyon-Caen A. (2002). *Rapport de la commission de réflexion sur les études de Droit*. Paris : Ministère de l'Éducation nationale.
- Magin D. & Helmore P. (2001). Peer and teacher assessments of oral presentation skills : how reliable are they ? *Studies in Higher Education*, 26 (3), 287-298.
- Mignot-Gérard S. & Musselin C. (2002). L'offre de formation universitaire : à la recherche de nouvelles régulations. *Éducation et sociétés*, 8, 11-25.
- Montgomery D. (1995). Critical theory and practice in evaluation and assessment. In G. Gibbs (ed). *Improving student learning through assessment and evaluation*. Oxford : Oxford Brookes University, 88-105.

- Morley L., Leonard D. & David M. (2002). Variations in vivas : quality and equality in british PhD assessments. *Studies in higher education*, 27 (3), 263-274.
- Musselin C. (2001). *La longue marche des universités françaises*. Paris : PUF.
- Navarro G. (2002). *L'évaluation des étudiants à l'université*. Communication à la 6^e Biennale de l'éducation et de la formation, Paris (2-5 juillet).
- Neave G. (1996). L'enseignement supérieur en transition. *Gestion de l'enseignement supérieur*, 8 (3), pp. 17-27.
- Noizet G. & Caverni J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Norton L. & Brunas-Wagstaff J. (2000). *Students' perception of fairness of assessment*. Paper given at ILTAC 2000 (The Institute for Learning and Teaching in higher education Annual Conference, College of Ripon & York St-John, 27-29 june).
- O.C.D.E. (1998). *Redéfinir l'enseignement tertiaire*. Paris : O.C.D.E.
- O'Donnovan B., Price M. & Rust C. (2000). The student experience of criterion-referenced assessment. *Innovations in education and teaching international*, 38 (1), 74-85.
- Pair C. (2001). *Forces et faiblesses de l'évaluation du système éducatif en France. Rapport établi à la demande du Haut Conseil pour l'évaluation de l'école*. Paris : Haut Conseil pour l'évaluation de l'école.
- Pascarella E. & Terenzini P. (1991). *How college affects students. Findings and insights from twenty years of research*. San Francisco : Jossey-Bass Publishers.
- Poirier Ch. (ed) (2001). *L'organisation et la préparation des examens : Compte rendu des Rencontres de l'Agence de Modernisation des universités et des établissements*. Paris : Agence de Modernisation des universités et des établissements.
- Potterton V. & Parsons P. (1995). Qualitative changes in learning and teaching brought about by student achievement. In G.Gibbs (ed). *Improving student learning through assessment and evaluation*. Oxford : Oxford Brookes University, 56- 69.
- Pratt D. (1997). Reconceptualizing the evaluation of teaching in higher education. *Higher education*, 34 (1), 23-44.
- Ramsden P. (1988). *Improving learning*. London : Kogan Page.
- Renaut A. (1995). *Les révolutions de l'université*. Paris : Calman-Lévy.
- Renaut A. (2002). *Que faire des universités ?* Paris : Bayard.
- Reynolds M. & Trehan K. (2000). Assessment : a critical perspective. *Studies in Higher Education*, 25 (3), 267-278.
- Romainville M. (1996). Enseignement et recherche : le couple maudit de l'université. *Gestion de l'enseignement supérieur*, 8 (2), 151-160.
- Romainville M. (2000). *L'échec dans l'université de masse*. Paris : l'Harmattan.
- Samuelowicz K. & Bain J. (2002). Identifying academics' orientation to assessement practice. *Higher education*, 43, 173-201.
- Scouller K. (1998). The influence of assessment method on students' learning approaches. *Higher Education*, 35 (4), 453-472.
- Segers M. & Dochy F. (2001). New assessment forms in problem-based learning. *Studies in Higher Education*, 26 (3), 327-343.
- Shattock M. (1999). L'impact du rapport Dearing sur l'enseignement supérieur au Royaume-Uni. *Gestion de l'enseignement supérieur*, 4, 7-18.
- Shevlin M., Banyard P., Davies M. & Griffiths M. (2000). The validity of student evaluation of teaching in higher education. *Assessment & Evaluation in Higher Education*, 25 (4), 397-405.
- Stray C. (2001). The shift from oral to written examinations : Cambridge and Oxford 1700-1900. *Assessment in Education*, 8 (1), 33-50.

- Tan C. (1992). An evaluation of the use of continuous assessment in the teaching of physiology. *Higher education*, 23, 255-272.
- Tynjälä P. (1998). Traditional studying for examination versus learning tasks : do learning outcomes differ ? *Studies in Higher Education*, 23 (2), 173-189.
- Wankat P. & Oreovicz F. (2002). Keeping students honest. *Prism (revue of American Society for Engineering Education)*, 12 (1) (<http://www.asee.org/prism/sept02/default.htm>)
- Warren Piper D. (1994). *Are professors professional ? The Organisation of University Examinations*. London : Jessica Kingsley Publishers.
- Webster F., Pepper D. & Jenkins A. (2000). Assessing the undergraduate dissertation. *Assessment and evaluation in Higher education*, 25 (1), 71-80.
- Wisker G. (1997). Assessing for learning in english studies some innovative practices. *Teaching in Higher Education*, 2 (2), 123-139.
- Woolf H. & Cooper A. (1999). Benchmarking academic standards in history. *Quality in higher education*, 5 (2), 145-154.
- York M., Bridges P. & Woolf H. (2000). Mark distributions and marking practices in UK higher education. *Active learning in higher education*, 1 (1), 7-27.

7.1 UNE ÉTUDE DE CAS : L'ÉVALUATION DES ACQUIS DANS DEUX DEUG D'HISTOIRE

Un cadre légal national peu contraignant...

Sur le plan national, le programme du DEUG histoire est défini par l'arrêté du 30 avril 1997 relatif au diplôme d'études universitaires générales Sciences humaines et sociales, aux licences et aux maîtrises du secteur Sciences humaines et sociales. Cet arrêté précise les objectifs généraux des DEUG de ce secteur et détermine les mentions (dont celle d'histoire). Il trace ensuite les grandes lignes du programme de DEUG en fixant la durée minimale des enseignements et la structure des deux semestres de la première année en six unités. Il fixe aussi la répartition entre les enseignements fondamentaux et les enseignements de découverte et de méthodologie ainsi qu'entre les cours magistraux et les autres activités d'enseignement. Pour chaque mention, l'arrêté fournit une liste des principaux domaines sur lesquels portent les enseignements fondamentaux. Pour l'histoire, cette description est minimaliste, puisqu'elle se contente d'évoquer les grandes périodes « histoire ancienne, histoire médiévale, histoire moderne, histoire contemporaine » ainsi que les « méthodes et techniques de l'histoire ». Dans le cadre de la large autonomie accordée aux établissements, le décret stipule enfin que chaque « *établissement définit et organise les différents cursus proposés aux étudiants.* »

Dans cet arrêté, aucune précision n'est apportée quant aux modalités de contrôle des connaissances. Bien sûr, les textes légaux généraux évoqués au point 3.2 encadrent cet aspect particulier. Mais, comme on l'a vu, ils ne concernent que les modalités externes des examens et ne touchent pratiquement pas au contenu de ceux-ci.

Des cursus aux couleurs locales...

Une première différence importante apparaît à la lecture des deux programmes. Dans l'université A, la première année est commune à deux mentions (histoire et sociologie). En deuxième année, l'étudiant peut, soit choisir une de ces mentions, soit opter pour les deux mentions et obtenir ainsi un double DEUG. Cette structure particulière a été conçue dans une triple perspective pédagogique.

- Un objectif d'interdisciplinarité : il s'agit de faire percevoir aux étudiants comment les deux disciplines s'articulent pour fournir des explications complémentaires des phénomènes sociaux. Cet objectif justifie aussi la création d'une unité d'enseignement interdisciplinaire autour d'un thème transversal, approché d'un double point de vue.
- Un objectif de rupture : cette structure des études a également été choisie pour provoquer une rupture dans les conceptions classiques qu'ont les étudiants de l'histoire à leur sortie du secondaire (histoire événementielle et politique) et pour les sensibiliser à d'autres dimensions de la discipline (histoire religieuse et des mentalités, par exemple).
- Un objectif d'orientation progressive des étudiants, ceux-ci choisissant leur mention définitive en connaissance de cause.

L'université B, tout en restant conforme aux mêmes prescriptions légales, a opté pour une autre perspective pédagogique. L'objectif est ici d'offrir aux étudiants des parcours différents selon leur projet professionnel. Ainsi, après un premier semestre plus « classique » que dans l'université A (l'unité d'enseignements fondamentaux est ici constituée d'une unité d'histoire du monde moderne et contemporain), le second semestre offre aux étudiants la possibilité de choisir un parcours géographie, de lettres ou de droit et sciences économiques.

L'objectif de cette étude de cas n'est pas de comparer systématiquement les programmes de ces deux DEUG, même si des divergences de cursus ont bien évidemment des conséquences importantes

en termes d'évaluation des acquis. Un travail minutieux de ce type serait pourtant très instructif⁴². Relevons simplement que les options pédagogiques prises dans ces deux établissements, parfaitement respectables les unes comme les autres et sans doute enracinées dans l'histoire des deux équipes d'enseignants, aboutissent à ce que leurs étudiants peuvent obtenir un même diplôme en ayant suivi des enseignements différents et donc en ayant sans doute développé des connaissances et des compétences différentes. Quelques exemples. Le jeu des options et des parcours peut aboutir à ce qu'un étudiant obtienne son DEUG dans l'université B sans avoir été initié à la sociologie alors que cette discipline est un axe prioritaire du cursus de l'université A. Même dans le cas d'enseignements fondamentaux similaires, les guides des études montrent que la coloration imposée par chaque enseignant peut amener à ce que les étudiants des deux sites soient, en définitive, confrontés à deux contenus de cours assez différents. Ainsi, une unité d'enseignements fondamentaux d'histoire contemporaine est consacrée, dans l'université A, à l'histoire large de l'Europe contemporaine centrée sur le mouvement des nationalités, alors que le cours correspondant de l'université B examine le cas particulier de la France, mais en le traitant selon les points de vue économique, social, culturel et religieux. Les orientations bibliographiques fournies aux étudiants dans les guides des études sont d'ailleurs différentes.

Les pratiques d'évaluation des acquis

Les modalités générales d'évaluation

Les modalités générales du contrôle des connaissances sont assez proches dans les deux DEUG, compte tenu sans doute des règles qui régissent cette matière (cf. 3.2). De plus, même des règles définies localement, conformément à l'autonomie pédagogique des universités qui laisse ces dernières mettre en œuvre concrètement les réglementations générales, sont fort semblables. Ainsi, dans les deux institutions, les modalités particulières à cette mention votées par le CEVU puis le CA stipulent que 50% de l'évaluation se réalisent par contrôle continu et 50% en contrôle terminal, avec les deux mêmes exceptions :

- l'unité de méthodologie (uniquement en contrôle continu) ;
- certains enseignements de découverte ou complémentaires qui ne se donnent qu'en cours magistral et ne connaissent dès lors qu'un contrôle terminal.

Des règles fort proches président aussi à la certification : réussite générale (avec une moyenne à 10), acquisition d'une unité capitalisable et compensation. Des divergences mineures apparaissent néanmoins. Ainsi, en ce qui concerne les coefficients, l'unité interdisciplinaire de l'université A se voit attribuer un coefficient de 2, alors que le coefficient maximum de l'université B est de 1,5. Plus important, les épreuves de rattrapage consistent pour l'essentiel en des examens oraux dans l'université B, alors que les deux formes d'examen (oral et écrit) doivent être proposées aux étudiants dans l'université A.

Les épreuves de contrôle continu

Les épreuves de contrôle continu apparaissent comme plus hétérogènes que les épreuves de contrôle terminal (cf. ci-dessous) et plusieurs enseignants se montrent relativement insatisfaits des pratiques actuelles. Les problèmes suivants sont évoqués, sur les deux sites.

⁴² Cependant, il ne serait guère aisé, compte tenu de la tradition universitaire de description elliptique des programmes de cours... Ainsi, il est parfois très difficile de se faire une idée du contenu précis des cours, même sur la base des guides d'études de chaque filière. Par exemple, les descriptions des unités de méthodologie ne fournissent pas beaucoup plus qu'une explicitation de l'objectif général poursuivi : « accéder à une démarche personnelle, indispensable à la réussite » et « aider les étudiants à s'adapter, au plus vite, aux exigences du supérieur ». Dans les faits, ces unités peuvent s'avérer exercer des compétences assez différentes selon l'importance accordée à la méthodologie générale du travail universitaire ou aux « méthodes propres à nos matières », à la compréhension ou à la production, etc.

- La difficulté d'évaluer les exposés oraux des étudiants. L'exposé oral est en effet assez répandu en contrôle continu et semble soulever des problèmes particuliers d'évaluation : comment se faire une idée exacte et précise de ce que maîtrise l'étudiant, alors que ce dernier conçoit l'oral comme « *un exercice d'érudition où il importe d'en montrer un maximum* » ? Comment s'assurer du caractère personnel de la préparation ? Comment évaluer en direct la qualité du raisonnement et de l'argumentation de l'étudiant, ces compétences se mesurant plus aisément dans un écrit sur lequel on peut revenir ?
- La machinerie complexe et lourde du contrôle continu : « *on a à peine commencé à faire cours que l'on évalue déjà* ». Les exigences du contrôle continu, surtout étalées sur un semestre, apparaissent assez lourdes : la note que l'étudiant obtient est parfois issue de six notes partielles, voire davantage. Plusieurs conséquences dommageables sont évoquées.
 - La lourdeur des corrections pour les enseignants.
 - Le risque d'évaluer trop tôt, c'est-à-dire avant que des apprentissages significatifs n'aient pu avoir lieu.
 - Le risque de passer en définitive plus de temps à évaluer qu'à enseigner⁴³ : par exemple, si l'évaluation porte sur un exposé oral et que le groupe d'étudiants est grand, l'essentiel du TD passe à une succession d'exposés, qui tournent chacun à une confrontation entre l'évaluateur et l'évalué, les autres étudiants ne se sentant pas concernés.
 - L'installation ou, du moins, l'encouragement d'un certain instrumentalisme de l'étude chez les étudiants : ils ne travaillent que s'ils ont une note et, dès qu'ils travaillent, il leur faut une note. Cette attitude empêche de recourir à des évaluations qui ne seraient que formatives : il semble, par exemple, impossible de leur demander un travail et de ne noter que certains d'entre eux, tirés au sort. Il est tout aussi difficile de multiplier les exercices de productions écrites, car les étudiants exigent alors qu'ils soient tous notés et l'enseignant se retrouve face à des centaines de copies...
 - La tendance à la notation centrale : la multiplication des notes portant sur des exercices divers (et parfois difficiles à noter en eux-mêmes, comme l'oral) aboutit à un lissage des performances des étudiants et à un resserrement des notes autour de la moyenne : « *le contrôle continu permet à certains étudiants de réussir parce qu'ils y accumulent des points sur des épreuves limitées.* »

Toutefois, les enseignants estiment que le contrôle continu est aussi une manière d'atténuer la rupture entre le secondaire et l'université. Les premières notes servent d'avertissement et les premiers feedbacks permettent à certains étudiants de corriger le tir : « *le contrôle continu, c'est lourd, mais si on en faisait moins, ce serait au détriment de l'étudiant* ».

Quand on analyse en détail le contenu précis des épreuves de contrôle continu des différentes unités, on peut faire l'hypothèse d'une grande variété probable de compétences mesurées en fonction du contenu et des méthodes des différents cours et TD eux-mêmes, ainsi qu'en fonction de la diversité des dispositifs concrets d'élaboration de la note de contrôle terminal :

- le nombre de notes intermédiaires ;
- l'importance de l'oral et le type d'oral (sujet imposé ou non) ;
- la présence ou non de partiels ;
- le type de tâches proposées : questions de restitution courte et ponctuelle ; exercices de chronologie ; compte rendu de lectures obligatoires ; rédaction de plans ; exercices d'écriture

⁴³ même si, comme le note malicieusement un enseignant, cette répartition du temps convient à certains enseignants désireux de n'investir qu'un minimum dans leur enseignement, au profit de leurs activités de recherche...

propres à la discipline, comme le traitement précis d'un sujet ou la rédaction d'une bibliographie,
...

De plus, dans leurs tentatives de concilier les avantages du contrôle continu avec la poursuite des activités d'enseignement, certains enseignants se voient contraints d'opter pour des formules d'évaluation qui risquent de poser des problèmes de validité et d'équité. Par exemple, les étudiants qui ne présentent pas l'oral au sein de tel TD remettent un écrit « *qui compte comme un oral* » et qui sera donc corrigé, plus ou moins, comme tel.

Les épreuves de contrôle terminal

Une plus grande homogénéité préside aux épreuves de contrôle terminal, du moins dans les enseignements fondamentaux. En effet, deux exercices « imposés » prédominent : la dissertation et le commentaire, au choix de l'étudiant. Ces exercices ont été en quelque sorte « canonisés » par les concours d'accès aux professions enseignantes (Capes et agrégation). Ils se sont donc imposés comme partie prenante et « naturelle » de l'évaluation qui sera celle des étudiants tout au long de leurs études, jusqu'à ces concours. Les enseignants interrogés se plient volontiers à ce moule, dans la mesure où, outre que ces épreuves constituent des passages obligés pour préparer les étudiants à ces concours, elles sont aussi considérées comme les exercices les plus formateurs de la discipline. Il y a donc un accord assez large sur le type d'épreuves à promouvoir, dès la première année du premier cycle, même si certains regrettent leur caractère parfois un peu formel (par exemple, les règles canoniques de l'introduction dans la dissertation) et la réduction de la variété des exercices qui en découle.

Les critères et les exigences

Toutefois, l'homogénéité du type d'épreuves ne garantit pas encore l'homogénéité des critères et des exigences. Les enseignants notent d'abord que, même s'il existe un accord général sur les exigences, des divergences peuvent apparaître en regard de la discipline ou de la spécialité à l'intérieur d'une même discipline. Ainsi, la capacité d'analyse de l'étudiant prendra une place plus importante pour le médiéviste, compte tenu de la place qu'occupe l'analyse des documents dans sa discipline.

Les procédures de correction font l'objet, sur les deux sites, d'un travail de collaboration entre les enseignants d'une même discipline qui préparent et homogénéisent l'épreuve⁴⁴. Il faut vraiment entrer dans le détail des procédures de correction et des critères pris en compte pour pouvoir observer des différences entre les deux équipes. Prenons un exemple qui a trait aux critères de correction des copies.

- Dans l'université B, les enseignants se mettent d'accord non seulement sur le sujet, mais aussi sur un barème de correction commun assez précis, conçu en termes de contenu de réponse attendu. Ainsi, les différents correcteurs ont à leur disposition un corrigé détaillé, d'une page environ, précisant les thèmes qui doivent figurer dans la réponse de l'étudiant ainsi qu'une ventilation des points correspondant à chacun de ces thèmes, fournie à titre indicatif. Les enseignants insistent sur le caractère souple du corrigé et du barème, ceux-ci ayant été conçus en « *fonction d'une copie idéale en tête* » : il s'agit dès lors de les confronter aux copies réelles, telles qu'elles ont été produites par les étudiants. Au cours de l'entretien et conformément au type de barème produit, les enseignants estiment qu'en première année de DEUG une restitution intelligente de points de matières glanés dans différentes parties du cours suffit à la réussite de l'épreuve. Les méthodes sont bien sûr importantes et certains « *incidents critiques* » pèsent lourd : un « hors sujet », une mauvaise valorisation des connaissances, une insuffisante maîtrise de l'écrit. Mais, comme le montre la conception du corrigé, l'essentiel des exigences a trait à la présence d'un nombre de thèmes incontournables dans la copie de l'étudiant, correctement explicités et exprimés.

⁴⁴ Il faut cependant noter que, dans chacune des deux universités, suite à notre demande de rencontrer des enseignants de la filière pour discuter de problèmes pédagogiques, nous avons été aiguillés vers des maîtres de conférences ou des Prags et qu'il n'est dès lors pas sûr que les préoccupations pédagogiques mises en avant dans cette étude de cas soient partagées par l'ensemble du corps des enseignants-chercheurs...

- Dans l'université A, les enseignants ont renoncé à utiliser des grilles explicites de correction, même s'ils fixent aussi les grands « passages obligés » et si une réunion de concertation peut avoir lieu a posteriori, si les différences de moyenne entre correcteurs sont trop importantes. En ce qui concerne les exigences, les enseignants estiment que « *les connaissances viennent au service d'une problématique* » et que les connaissances de l'étudiant ne sont donc pas valorisées en tant que telles : « *si l'étudiant accumule beaucoup de connaissances correctes mais sans articulation, il n'a pas la moyenne* ». Un autre signe du primat du raisonnement et de l'argumentation sur les connaissances réside dans le fait que les grilles d'aide à l'évaluation qui avaient été conçues par l'équipe étaient rédigées en termes de « savoir-faire à mettre en œuvre » dans le commentaire de texte et dans la dissertation.

Bien sûr, cette différenciation des sites selon l'importance accordée aux connaissances ou aux compétences est trop dichotomique et réductrice. De nombreuses nuances seraient à apporter, ne fût-ce que par la diversité des conceptions exprimées par les enseignants d'une même équipe. Ainsi, les enseignants de l'université A fixent aussi des « passages obligés » en termes de contenus et les enseignants de l'université B notent que des incidents critiques ayant trait au savoir-faire pèsent dans la balance. Il ne s'agit donc que d'une coloration particulière apportée par chaque équipe à leurs exigences. Mais, dans le cadre de ce rapport sur l'évaluation des acquis, l'essentiel est de montrer comment une simple coloration spécifique des exigences, même dans le cas d'épreuves canoniques, porte sans doute à conséquence au point de vue des acquis qui sont effectivement sanctionnés par le diplôme.

7.2 PERCEPTIONS DES ACQUIS CHEZ LES ÉTUDIANTS (D'APRÈS DREW, 1998)

Nbre d'étudiants sur 313 questionnaires	Nbre de groupes sur 14 entretiens	Acquis perçus
271	14	Gestion de soi <ul style="list-style-type: none"> ▪ Organisation personnelle (gérer et organiser son temps, respecter des échéances, se fixer des priorités, organiser son travail de manière méthodique et fiable, équilibrer travail et loisirs) ▪ Indépendance (confiance en soi, autodiscipline, motivation intrinsèque, responsabilité, opérer des choix personnels) ▪ Capacité à faire face (affronter des situations nouvelles et perturbantes, gérer l'urgence, le stress, perséverer) ▪ Gestion financière
256	14	Compétences relationnelles <ul style="list-style-type: none"> ▪ Compétences sociales (assertivité, tolérance à la critique, capacité de persuasion et d'influence, de négociation, diplomatie, écoute, patience, capacité au compromis, tolérance et respect des différences) ▪ Capacité à entrer en relation (nouer des amitiés, fréquenter des personnes d'horizons variés, répondre aux sollicitations d'autrui) ▪ Capacité à travailler en groupe (coopérer, déléguer)
178	13	Compétences de communication <ul style="list-style-type: none"> ▪ Communication orale (prise de parole, exposé, participation à des discussions, capacité à poser des questions, à expliquer, à argumenter) ▪ Communication écrite (lettre, rapport, essai, cohérence textuelle, style et genre académique, syntaxe et orthographe) ▪ Présentation visuelle ▪ Écoute ▪ Présentation de soi (interview et CV) ▪ Communication non-verbale
140	12	Connaissance de son environnement <ul style="list-style-type: none"> ▪ Découverte de champs professionnels, des mondes économique, politique et social ▪ Élargissement de ses horizons
131	10	Connaissances (spécifiques et générales)
129	13	Confiance en soi
125	11	Conscience de soi et développement personnel <ul style="list-style-type: none"> ▪ Développement d'un projet professionnel ▪ Capacité de s'autoévaluer, de se critiquer ▪ Développement moral ▪ Apprendre à apprendre
115	11	Compétences informatiques
104	11	Changement d'attitudes

		<ul style="list-style-type: none"> ▪ Ouverture, attitude moins dogmatique, plus tolérante ▪ Attitude plus professionnelle ▪ Réalisme
103	12	Techniques d'étude et d'apprentissage <ul style="list-style-type: none"> ▪ Utilisation des bibliothèques et de ressources d'information ▪ Maîtrise des outils de recherche d'information
82	7	Habilités pratiques
40	4	Habilités intellectuelles Pensée critique, analyse, synthèse, application, concentration, pensée complexe...
29	8	Résolution de problèmes
29	4	Créativité

7.3 LE SYSTÈME ANGLAIS D'EXAMINATEURS EXTERNES

Chaque institution d'enseignement supérieur britannique est tenue, selon un code de « bonne conduite » interne à l'association des directions des établissements d'enseignement supérieur, de faire appel à des examinateurs externes qu'elle rémunère elle-même. Ces examinateurs externes dépendent directement de la direction de l'institution. Il s'agit d'enseignants-chercheurs extérieurs, experts et indépendants, provenant d'autres institutions universitaires ou du monde professionnel. Ils sont engagés à ce titre pour une période de trois à cinq ans. En étant associés au travail des jurys d'examen, leur tâche consiste à rendre des avis sur les points suivants :

- les critères d'évaluation sont-ils appropriés aux standards nationaux et les étudiants sont-ils correctement et équitablement jugés par rapport à ces critères ?
- les critères sont-ils comparables à ceux qui sont utilisés dans d'autres établissements pour la même filière ?
- les pratiques d'évaluation des acquis sont-elles valables et sont-elles conduites de manière juste et fiable ?

Sur la base de ces principes, les établissements jouissent d'une grande liberté quant à l'organisation concrète du travail de ces examinateurs externes.

- Dans quelle mesure participent-ils à la composition des épreuves (d'une simple approbation à la possibilité de suggérer ou de réaliser des modifications) ?
- Sur quel matériel d'évaluation travaillent-ils (de l'ensemble des copies de tous les candidats jusqu'à un échantillon représentatif, proposé ou construit par eux, ou un échantillon de « cas limites ») ? En particulier, quel est leur rôle quant aux épreuves orales ?
- Quel est leur rôle en cas de désaccord concernant le résultat global d'un candidat (de la conciliation à l'imposition d'une décision) ?
- Font-ils des recommandations quant à des ajustements de notes ou de mentions, individuelles ou collectives ? Sur quelle base ?

À titre d'illustration, voici un court témoignage d'un professeur qui a eu l'occasion d'enseigner quelques années dans une prestigieuse université anglaise :

« L'examineur externe intervient tout au long du processus d'évaluation. Sa première tâche est de se prononcer sur les projets d'examens. Il lit les projets de questions, en rapport avec les objectifs et le polycopié du cours. S'il estime que l'examen cadre avec les objectifs annoncés et correspond aux standards en vigueur, il donne son imprimatur. Sinon, il suggère des amendements. La seconde tâche consiste à contrôler la cotation des examens. Typiquement, lui sont envoyées les copies de tous les étudiants qui ont soit échoué, soit réussi un excellent examen ; on leur envoie aussi un échantillon des copies « intermédiaires ». L'examineur externe se prononce sur la cohérence de la distribution des notes. Parfois, une demande expresse lui est faite quant à tel ou tel cas « limite ». La troisième tâche est d'assister aux réunions des jurys d'examens, avec la même voix que les membres internes. Au terme du processus, l'examineur externe rédige un rapport sur le ou les cours qu'il a supervisé(s) et sur l'ensemble de l'organisation de la session d'examens. Le travail de l'examineur externe est pris très au sérieux. Le travail est fait sans complaisance et avec professionnalisme, en dépit d'une rémunération assez symbolique. Les examinateurs externes sont bien accueillis et respectés, sans doute parce que tout le monde participe un peu au système ; certains de mes collègues jouaient ce rôle auprès d'autres universités. »

Bien sûr, des voix se sont élevées au Royaume-Uni pour dénoncer le risque que le système d'examineurs externes, même si ce contrôle externe et original de la qualité des pratiques

d'évaluation appartient à une longue tradition nationale⁴⁵, empiète trop sur l'autonomie des établissements (Shattock, 1999). Pour plus de détails, ce système est décrit au chapitre 2 de l'ouvrage de Warren Piper (1994) ainsi que sur le site de l'Agence pour l'assurance qualité dans le supérieur⁴⁶.

⁴⁵ On en trouve des traces dès le 16^e siècle...

⁴⁶ <http://www.qaa.ac.uk/public/COP/COPee/contents.htm>

1	INTRODUCTION	3
1.1	OBJECTIFS ET PLAN DU RAPPORT	3
1.1.1	<i>Que sait-on des acquis des étudiants ?</i>	3
1.1.2	<i>Pourquoi sait-on si peu de choses sur les acquis des étudiants ? Les pratiques d'évaluation à l'université.....</i>	3
1.1.3	<i>Pourquoi devrait-on en savoir davantage ? Faiblesses et lacunes de l'évaluation des acquis</i>	4
1.1.4	<i>Comment améliorer les pratiques d'évaluation des acquis ?</i>	4
1.2	MÉTHODOLOGIE.....	4
2	QUE SAIT-ON DES ACQUIS DES ÉTUDIANTS ?.....	7
2.1	DES INVENTAIRES DE COMPÉTENCES D'ENTRÉE	7
2.2	DES ÉTUDES SUR LA PERCEPTION QU'ONT LES ÉTUDIANTS ET LEURS EMPLOYEURS DES COMPÉTENCES ACQUISES.....	7
2.3	DES INVENTAIRES DE COMPÉTENCES MÉTHODOLOGIQUES	8
2.4	DES ÉTUDES SUR L'EFFET GÉNÉRAL DE L'ENSEIGNEMENT SUPÉRIEUR.....	9
2.4.1	<i>Les acquis de connaissances et de compétences.....</i>	9
2.4.2	<i>Les acquis de savoir-faire cognitifs et de développement intellectuel.....</i>	10
2.4.3	<i>Autres acquis</i>	11
3	POURQUOI SAIT-ON SI PEU DE CHOSES ? LES PRATIQUES D'ÉVALUATION DES ACQUIS À L'UNIVERSITÉ	13
3.1	UN ALLOURDISSEMENT DES TÂCHES D'ÉVALUATION, AYANT DES EFFETS NÉGATIFS SUR SA QUALITÉ	14
3.2	DES PRATIQUES D'ÉVALUATION PEU STANDARDISÉES.....	15
3.2.1	<i>Absence de standardisation des objectifs, des contenus et des méthodes.....</i>	18
3.2.2	<i>Une longue tradition de « liberté académique ».....</i>	20
3.2.3	<i>Une grande diversité de conceptions de l'évaluation.....</i>	21
3.2.4	<i>Une mixité de fonctions difficilement conciliables</i>	21
3.3	UNE ÉVALUATION NORMATIVE, QUI NE SE FONDE PAS SUR DES OBJECTIFS EXPLICITES DE FORMATION	22
3.3.1	<i>Des formations qui ne sont pas définies en termes de compétences à acquérir</i>	24
3.3.2	<i>Une absence de reconnaissance du caractère « professionnel » de l'acte d'évaluation.....</i>	26
3.3.3	<i>Une évaluation à fonction sélective et non d'inventaire.....</i>	28
3.4	UN MANQUE DE VALIDITÉ, DE FIDÉLITÉ ET DE FIABILITÉ.....	28
3.5	UNE ÉVALUATION QUI « PILOTE » LES ÉTUDIANTS.....	32
3.5.1	<i>L'évaluation par la restitution appelle l'étude en surface.....</i>	33
3.5.2	<i>L'évaluation par la restitution est peu compatible avec les plus hautes finalités de l'enseignement universitaire</i>	34
4	POURQUOI DEVRAIT-ON EN SAVOIR DAVANTAGE ? LIMITES ET LACUNES DES PRATIQUES ACTUELLES D'ÉVALUATION DES ACQUIS	37
4.1	LES RAPPORTS PROSPECTIFS SUR L'ENSEIGNEMENT SUPÉRIEUR	37
4.1.1	<i>La certification de la qualité des acquis.....</i>	37
4.1.2	<i>L'employabilité.....</i>	38
4.2	LES EFFETS DOMMAGEABLES	39
5	COMMENT AMÉLIORER LES PRATIQUES D'ÉVALUATION DES ACQUIS ?.....	43
5.1	LE DÉVELOPPEMENT D'ÉPREUVES STANDARDISÉES ET INDICATIVES D'ÉVALUATION DES ACQUIS	43
5.2	LA PROFESSIONNALISATION DE L'ÉVALUATION LOCALE DES ACQUIS.....	45
5.2.1	<i>Le développement d'une culture institutionnelle de la qualité autour de l'évaluation</i>	45

5.2.2	<i>La formation des enseignants-chercheurs à l'évaluation</i>	46
5.2.3	<i>La clarification du « contrat pédagogique »</i>	47
5.2.4	<i>Le développement de pratiques innovantes d'évaluation des acquis</i>	48
5.3	UNE PLUS GRANDE PRISE EN COMPTE DES PRATIQUES D'ÉVALUATION DES ACQUIS DANS LES PROCÉDURES DE RÉGULATION	49
6	RÉFÉRENCES	51
7	ANNEXES	57
7.1	UNE ÉTUDE DE CAS : L'ÉVALUATION DES ACQUIS DANS DEUX DEUG D'HISTOIRE	57
	<i>Un cadre légal national peu contraignant</i>	57
	<i>Des cursus aux couleurs locales</i>	57
	<i>Les pratiques d'évaluation des acquis</i>	58
7.2	PERCEPTIONS DES ACQUIS CHEZ LES ÉTUDIANTS (D'APRÈS DREW, 1998).....	62
7.3	LE SYSTÈME ANGLAIS D'EXAMINATEURS EXTERNES	64