

L'évaluation des apprentissages des étudiants en formation médicale initiale

Jean JOUQUAN*

Messages clés. L'évaluation des apprentissages est un processus qui consiste à recueillir des informations relatives à ces apprentissages, à les interpréter afin de porter un jugement de valeur, dans le but de prendre une décision. ● Compte tenu du rôle central de l'évaluation dans la dynamique enseignement-apprentissage, il importe fondamentalement de se préoccuper de la cohérence du dispositif d'évaluation avec les orientations conceptuelles du programme de formation. ● Les pratiques pédagogiques issues du modèle de « l'enseignant qui enseigne à l'étudiant » sont plus souvent associées à une approche évaluative qui privilégie la mesure des apprentissages à partir de comportements observables, dans des tâches standardisées ; les pratiques pédagogiques issues du modèle de « l'étudiant qui apprend individuellement, à partir des questions qu'il s'est posées, en interaction avec le professeur » sont plus souvent associées à une approche évaluative qui privilégie des sources d'informations plus éclectiques concernant les apprentissages, recueillies lors de tâches authentiques et complexes, intégrées aux activités d'enseignement et d'apprentissage. ● Plusieurs recommandations générales méritent d'être prises en compte lors du développement d'un dispositif d'évaluation ; anticiper les conséquences prévisibles du dispositif mis en oeuvre sur les apprentissages des étudiants ; délimiter ce qui doit être évalué en raison de sa pertinence par rapport aux compétences visées plutôt qu'en raison de sa facilité à être mesuré ; sélectionner des modalités d'évaluation qui renseignent valablement sur le niveau de performance réellement attendu des étudiants ; exposer les étudiants à un nombre et à une variété suffisants de situations d'évaluation ; développer des dispositifs d'évaluation utilisant des méthodes et des instruments que l'on maîtrise, dont on connaît les forces et les limites, compatibles avec les allocations en ressources du programme de formation et acceptés par les acteurs concernés ; rendre explicite aux étudiants, dès le début de la séquence d'enseignement et d'apprentissage, la nature des tâches d'évaluation et les critères pris en compte pour la correction et/ou pour la détermination de la réussite.

Mots clés Evaluation des apprentissages ; paradigme pédagogique ; docimologie ; évaluation authentique.

Key messages. Learning assessment is a process which consists in the collection of data relating to student learning and their interpretation in order to make value-based decisions. ● Considering the central role of assessment in the teaching and learning process, it is of basic importance to be concerned by the coherence of the evaluation process with the conceptual orientations of the program. ● On one hand, strong convergences exist thereby between the teacher-centered educational paradigm, behavioral viewpoint of learning and the docimological assessment movement; on the other hand, strong convergences exist between the learner-centered educational paradigm, constructivist viewpoint of learning and the authentic assessment movement. ● Several general guidelines should be taken into account during the development of an assessment plan: anticipate the foreseeable consequences of the evaluation implemented on student learning; choose learning outcomes more in accordance with their relevance with expected competences rather than because of their easiness to be measured; select methods and tools of assessment which give a valid information on the level of performance actually expected by the students; expose students to a sufficiently wide and eclectic range of assessment situations; implement assessment programs that use well mastered methods and tools, with known strengths and limits, compatible with the resources of the program and which are properly understood and accepted by all the actors concerned; provide explicit informations to the students regarding the nature and the features of the assessment procedures and criteria that will be taken into account for evaluation decisions.

Key words learning assessment ; educational paradigm ; docimology ; authentic assessment.

Pédagogie Médicale 2002 ; 3 : 38-52

*Bureau de Pédagogie Médicale (Faculté de Médecine) et Unité de Recherche Action en Formation de Formateurs (URAFF, Université de Bretagne Occidentale) - Brest - France.

Correspondance : Pr. J. JOUQUAN - CHU de la Cavale Blanche - 29609 BREST cedex - France - Tél. : (33) 02 98 34 71 91
Télécopieur : (33) 02 98 34 71 93 - Mailto:jean.jouquan@chu-brest.fr

Introduction

L'évaluation est fondamentalement une démarche de recueil d'informations concernant une activité ou son résultat, conduisant à un jugement et à une prise de décision¹. En éducation, un tel processus peut concerner schématiquement les établissements et les institutions de formation, les programmes et les dispositifs de formation, les enseignants ou les étudiants.

En tant que concept, l'évaluation des étudiants -ou plus précisément des apprentissages des étudiants- est aujourd'hui tenue comme l'un des éléments essentiels du processus enseignement-apprentissage ; à ce titre, les cadres théoriques qui l'inspirent rendent compte des divers courants de la pédagogie. En tant que pratique, l'évaluation des apprentissages des individus engagés dans un programme de formation recouvre des procédures extrêmement hétérogènes, dont les buts sont, en outre, variables. De ce fait, le champ de l'évaluation, en pédagogie, est souvent perçu par les enseignants comme extrêmement complexe. Le risque est alors grand que certains d'entre eux n'accordent pas à l'évaluation autant de soin qu'aux autres actes qui concourent à leur rôle de formateur, allant parfois jusqu'à en sous-traiter la mise en œuvre ; d'autres au contraire, confondant buts et moyens, sont tentés d'élaborer des outils et de construire des dispositifs d'évaluation aux proportions démesurées.

Le domaine de l'éducation médicale, et spécifiquement celui de la formation médicale initiale, n'échappe pas à ces écueils ; ceux-ci sont encore renforcés par la longueur des études, par la multiplicité des formateurs et des milieux de formation concernés, par la nature et la complexité des compétences visées en fin de cursus et enfin par les attentes particulièrement fortes de la société à l'égard de la profession médicale, qui confèrent au mandat des évaluateurs des contraintes et des exigences particulières. Une abondante littérature internationale rend compte du foisonnement des recherches et des pratiques concernant la démarche évaluative en formation médicale. Cette littérature n'est pas toujours d'accès facile car elle se réfère plus ou moins explicitement à des cadres conceptuels multiples, eux-mêmes plus ou moins bien stabilisés ; pour décrire les formats et les pratiques d'évaluation développés, elle utilise, en outre, une terminologie protéiforme, majoritairement développée en anglais et dont les équivalents français sont parfois incertains voire inconnus.

Ainsi, la présente contribution s'est donnée comme

objectif d'apporter à un lecteur curieux -mais pas nécessairement spécialiste de cette littérature- quelques points de repère formulés de manière didactique -et donc nécessairement réductrice-, concernant la problématique de l'évaluation des apprentissages en formation médicale initiale. Nous développeront successivement quelques définitions, les différentes approches conceptuelles, certains aspects opérationnels et quelques perspectives que l'on peut entrevoir en ce domaine.

1. Définitions

Définitions générales

L'évaluation des apprentissages est un processus plus ou moins formel qui comporte trois étapes¹ : a) le recueil d'informations relatives aux apprentissages de toute nature réalisés par un étudiant pendant puis à l'issue d'une séquence de formation ; b) l'interprétation de ces informations afin d'être en mesure de porter un jugement de valeur concernant ces apprentissages ; c) la prise d'une décision, de nature variable, fondée sur le jugement porté.

Il est fondamental de bien comprendre que la démarche évaluative recouvre l'ensemble de ces trois étapes ; chacune d'entre elles ne prend son sens que par rapport aux autres et aucune n'est redondante, ce qui signifie que l'on ne peut s'affranchir ni de l'une, ni de l'autre. Plusieurs conceptions erronées sont pourtant répandues à cet égard. L'une d'entre elles consiste à confondre évaluation et mesure. La mesure, à l'aide de tests par exemple, ne constitue que l'un des moyens possibles d'obtenir des informations relatives à des apprentissages ; elle fournit des informations traduites en variables quantitatives¹. En ce sens, la mesure est une activité beaucoup plus restreinte que l'évaluation, à la fois parce que celle-ci exige souvent que l'on ait aussi accès à des informations qualitatives de nature plus complexe, difficiles à quantifier et aussi parce que, dans une optique d'évaluation et même lorsqu'elles sont issues d'une mesure, les informations doivent être interprétées avant de conduire à une décision. A cet égard, la mesure est elle-même souvent confondue avec la note. *Stricto sensu*¹, le résultat d'une mesure est un score, c'est-à-dire un nombre exprimant le résultat d'un test ou la somme des résultats obtenus à plusieurs tests. La note, quant à elle, est une valeur numérique

Références

(8 sur 20, 80 %,...) ou encore une cote (lettre ou symbole) qui traduit l'appréciation quantitative et/ou qualitative portée, selon un barème, sur les apprentissages et le travail d'un étudiant ; c'est une manière codifiée de communiquer le résultat d'une évaluation à un étudiant et de le rendre public. Les confusions mentionnées résultent en fait de l'influence extrêmement forte, en matière d'évaluation, du courant fondé sur une théorie de la mesure et que l'on peut assimiler au courant docimologique. Nous en développerons les fondements plus loin en le situant par rapport à d'autres courants développés en évaluation, notamment celui de l'évaluation dite « authentique » Il nous paraissait cependant primordial d'indiquer d'emblée qu'une séquence consistant à énoncer : « cet étudiant a obtenu 11/20 à ses tests (mesure) ; donc il a fait des apprentissages satisfaisants (jugement) ; en conséquence, il est reçu à son examen (décision) » constitue une démarche d'évaluation extrêmement réductrice, incomplète, de nature comptable, au cours de laquelle la seule mesure tient lieu tout à la fois de procédure de recueil de l'information, de jugement et de décision.

Définitions spécifiques

Évaluation sommative et évaluation formative

Il est devenu classique en pédagogie de distinguer d'un côté l'évaluation sommative et, de l'autre, l'évaluation formative des apprentissages. Malgré son évidence de premier degré et son apparente clarté didactique, la validité réelle d'une telle distinction repose en réalité sur des choix théoriques et conceptuels assez complexes concernant la dynamique enseignement-apprentissage^{2, 3, 4}. Le développement de ceux-ci dépasserait les limites imparties à ce texte. Parce qu'elle est consacrée par l'usage et qu'elle est commode pour décrire les différentes pratiques évaluatives, cette typologie de l'évaluation reste aujourd'hui incontournable et la littérature en éducation médicale y fait régulièrement référence.

L'évaluation sommative des apprentissages est donc celle qui intervient à la fin d'une séquence d'enseignement afin de juger du degré et de la valeur des apprentissages réalisés par l'étudiant. Elle est le plus souvent utilisée pour décider de la réussite ou de l'échec de l'étudiant à tout ou partie d'un programme d'études, son passage en année supérieure, l'octroi d'un diplôme, d'une autorisation d'exercice ou d'une certification.

De ce fait, elle est souvent identifiée à la fonction administrative de l'évaluation et est parfois qualifiée de sanctionnante.

Par contraste, l'évaluation formative des apprentissages a une fonction diagnostique et vise à réguler les apprentissages. Il s'agit, à des moments variables ou de manière continue, de donner à l'étudiant et à l'enseignant des informations objectives sur la nature et la valeur des apprentissages réalisés, afin que l'un et l'autre ajustent et optimisent, respectivement, leurs stratégies d'apprentissage ou leurs interventions pédagogiques. Elle est généralement assimilée à la fonction pédagogique de l'évaluation.

Interprétation normative et interprétation critériée

Pour fonder le jugement de valeur, étape centrale du processus évaluatif, les informations recueillies au sujet des apprentissages doivent être interprétées. Lorsque ces informations résultent d'une mesure, cette interprétation peut résulter d'une approche normative ou d'une approche critériée.

Lors d'une mesure à interprétation normative, le résultat de l'étudiant est comparé à celui des autres étudiants de son groupe ou à celui d'un groupe de référence ayant subi les mêmes épreuves. Les modalités de ces épreuves doivent notamment posséder la capacité de différencier les étudiants « faibles » par rapport aux étudiants « forts », l'information sur les apprentissages réalisés étant alors moins importante que la position de l'étudiant par rapport aux étudiants de son groupe.

Lors d'une mesure à interprétation critériée, le résultat de l'étudiant est comparé à un critère prédéterminé, en termes de performances par rapport à une tâche ou un panel défini de tâches. Il s'agit d'apprécier le niveau d'apprentissage d'un étudiant par rapport aux objectifs d'un programme de formation et non par référence à un groupe. Lorsqu'une telle interprétation doit fonder une évaluation sommative, la formulation de l'objectif d'apprentissage comporte la description d'un seuil de réussite minimal.

Par nature, les informations de nature quantitative peuvent se prêter autant à une interprétation normative qu'à une interprétation critériée, ces deux modalités ne s'opposant pas nécessairement mais pouvant être complémentaires en fonction du but poursuivi. Lorsque les informations sont de nature qualitative, elles ne font en règle l'objet que d'une interprétation critériée. L'interprétation critériée est surtout adaptée

aux décisions concernant l'apprentissage de l'étudiant, son cheminement individuel et, le cas échéant, sa certification tandis que l'interprétation normative contribue surtout à des décisions concernant l'apprentissage d'un groupe en général, le classement ou certaines procédures de certification fondées sur une sélection.

2. Les différentes approches conceptuelles de l'évaluation des apprentissages

Une perspective historique des courants théoriques de l'évaluation des apprentissages

Les pratiques évaluatives sont d'une extraordinaire diversité quant aux formats qu'elles adoptent et aux outils qu'elles utilisent et elles répondent en outre à des buts non univoques. Cette hétérogénéité n'est cependant pas le résultat d'une génération « spontanée » totalement débridée. Même si l'on n'en a pas toujours explicitement conscience, chaque pratique évaluative est en réalité fortement influencée par un ou plusieurs courants conceptuels qui en constituent les fondements et qui en déterminent à la fois les forces et les limites. Ces différentes influences ne sont pas nécessairement exclusives l'une de l'autre et plusieurs pratiques évaluatives empruntent à l'une et à l'autre. Leur développement répond néanmoins globalement à une chronologie historique ; il est possible de résumer les grandes lignes de celle-ci en s'appuyant sur la typologie reprise par Louis⁵, qui distingue trois approches.

Une approche centrée sur les différences individuelles

Un premier courant, développé dans les années cinquante et soixante, postule globalement que les apprentissages d'un étudiant résultent d'habiletés individuelles constitutives de son intelligence, celle-ci étant perçue dans une optique néo-darwinienne comme une caractéristique plus ou moins fixe, résultat du processus général de l'évolution. Le but de l'évaluation est alors de catégoriser les individus en fonction de ces aptitudes, en utilisant une batterie de tests qui explorent un répertoire de traits cognitifs unidimensionnels. L'interprétation des informations est normative puisque, dans cette perspective, la position des individus par rapport aux autres est fondamentale.

Cette approche de l'évaluation, que l'on peut qualifier d'endogène, s'est fort logiquement appuyée sur le courant de la psychométrie en éducation, la somme des scores obtenus aux différents tests étant conçue comme un indicateur valide des apprentissages réalisés. Même s'il faut se garder de toute caricature, on peut reconnaître le caractère très prégnant de cette approche dans les pratiques évaluatives qui visent à organiser le cheminement professionnel des étudiants sur la base d'une sélection ; la généralisation des concours en France, par exemple, fondée sur une conception très « républicaine » du mérite et de l'équité, en est une illustration très emblématique et il n'est pas neutre que, situation à peu près unique dans le monde, la séquence indifférenciée des études de médecine dans ce pays (premier et deuxième cycles, équivalents des études « pré-graduées » nord-américaines) débute et finisse par un concours de sélection.

Une approche centrée sur les contenus externes à l'individu

Un deuxième courant a marqué fortement le développement des pratiques évaluatives des années 1970 aux années 1990. Il assimile globalement les apprentissages à la quantité du contenu d'un programme d'études que l'étudiant maîtrise à l'issue d'une séquence de formation. L'accent est ainsi mis sur le contenu externe à l'individu, ce qui autorise à qualifier cette approche d'exogène. Ce contenu est soigneusement traduit en objectifs d'apprentissages, le nombre d'objectifs atteints étant tenu pour constituer un indicateur satisfaisant des apprentissages des étudiants. Ce courant inspire très majoritairement les conceptions et les pratiques actuelles de l'évaluation en formation médicale et, pour de multiples raisons, il est prévisible qu'il restera pendant longtemps fortement influent.

Une approche centrée sur l'interaction de la personne et de l'environnement

Le troisième courant est en fait relativement récent puisqu'il ne se développe opérationnellement que depuis les années 1990, même si ses fondements théoriques, d'inspiration constructiviste, sont plus anciens. Il s'appuie essentiellement sur l'idée que les apprentissages sont des processus éminemment individuels, qui aboutissent à des modifications de la structure cognitive singulière de chaque individu apprenant. Il prend également en compte l'influence déterminante du

Références

contexte dans les activités de traitement de l'information et tente d'apporter des solutions à la nécessité de vérifier que les apprentissages sont transférables aux contextes prévisibles de réutilisation, en l'occurrence les situations de résolution de problèmes complexes de santé. Ce courant, parfois qualifié d'écologique, se recentre donc sur l'interaction entre l'apprenant et son environnement. Lorsqu'il concerne l'évaluation d'apprentissages non plus élémentaires mais celle d'apprentissages complexes, tels que ceux impliqués dans la construction de la compétence médicale, et contrairement au courant précédent, il postule que la résultante des divers apprentissages est d'une autre nature que la stricte somme des apprentissages élémentaires. De ce fait, il s'intéresse autant au processus d'apprentissage qu'à son résultat, retient le caractère pluridimensionnel des traits à évaluer et considère que l'évaluation doit s'attacher à apprécier un profil de fonctionnement cognitif de l'individu dans un contexte donné, plus ou moins contrôlé ou aléatoire.

Une perspective systémique des cadres conceptuels et des pratiques en évaluation

La perspective historique très brièvement brossée dans les lignes précédentes illustre bien que les influences conceptuelles exercées sur les pratiques évaluatives sont multiples. Des contraintes de tous ordres (social, institutionnel, professionnel, financier, logistique,...) expliquent, en outre, que les choix à faire par un enseignant, un groupe d'enseignants, un établissement ou une institution sont complexes. Il faut pourtant être conscient que les modalités d'évaluation mises en œuvre conditionnent fortement la qualité, la nature ainsi que le caractère réutilisable des apprentissages que les étudiants font. Adopter une perspective systémique en évaluation signifie que l'on se préoccupe fondamentalement des interactions que les pratiques évaluatives génèrent à l'égard des autres composantes de la dynamique enseignement-apprentissage et de son résultat, à l'intérieur d'un dispositif de formation.

Le rôle central de l'évaluation dans la dynamique enseignement-apprentissage

Pendant longtemps, le processus de planification pédagogique a été décrit en situant l'activité d'évaluation des apprentissages à la fin du cycle enseignement/apprentissage. Les procédures évaluatives étaient

souvent planifiées et mises en œuvre après les étapes d'identification des besoins, de formulation des objectifs et d'intervention pédagogique. Il s'agissait alors d'une démarche construite de façon très linéaire, même si l'on indiquait que le processus devait être indéfiniment réitéré, « en boucle » (la fameuse « spirale de l'éducation »).

On avait pourtant très tôt compris à quel point les modalités d'évaluation des apprentissages conditionnaient la nature et la qualité des apprentissages que les étudiants développaient. La locution anglophone « *evaluation drives curriculum* », qui a pris valeur d'aphorisme, traduit ainsi l'idée que les étudiants s'adaptent fondamentalement et stratégiquement à ce qui est attendu d'eux ; en conséquence, il est hautement probable que l'on a plus de chances d'influencer significativement leurs apprentissages en modifiant le dispositif d'évaluation sans changer les procédures d'intervention pédagogique qu'en modifiant celles-ci, même en profondeur, sans changer les procédures de l'évaluation.

De fait, l'évaluation doit être tenue pour centrale dans la dynamique enseignement-apprentissage. Elle conditionne, par exemple, fortement les orientations des apprentissages. On sait à cet égard que l'intention avec laquelle on apprend et la nature du traitement de l'information que l'on développe pour ce faire déterminent puissamment la qualité, la profondeur et le caractère réutilisable des apprentissages ; en d'autres termes, à partir d'un même corpus d'informations, on n'effectue pas du tout les mêmes apprentissages si l'on apprend dans le but d'être capable de réciter ces informations ou si l'on apprend dans le but d'être capable de les réutiliser pour comprendre, expliquer et résoudre un problème. De ce fait, les éléments pris en compte à l'étape de l'évaluation par les professeurs envoient aux étudiants des messages puissants concernant les aspects à privilégier dans leurs apprentissages. L'évaluation influence également fortement la motivation d'un étudiant à apprendre^{6, 7}, en lui fournissant des informations qui participent à son sentiment de compétence par rapport aux tâches d'apprentissage proposées. En l'aidant à en percevoir la valeur et à identifier le degré de contrôle qu'il possède à leur égard, elle peut favoriser l'engagement et la persévérance de l'étudiant.

La cohérence du dispositif d'évaluation avec les orientations conceptuelles du programme de formation

Aucune méthode ni aucun outil d'évaluation n'est bon ou mauvais en soi. Les méthodes et les outils sélectionnés doivent, en revanche, être fondamentalement cohérents avec les choix pédagogiques du programme de formation. Il s'agit, en d'autres termes, d'accroître la probabilité que les conséquences prévisibles des pratiques évaluatives à l'égard des apprentissages des étudiants rejoignent les intentions des enseignants et du programme. A cet égard, il faut considérer que de fortes convergences -et, dès lors, une solide cohérence- existent entre le paradigme d'enseignement, la perspective behavioriste de l'apprentissage et le courant docimologique en évaluation d'une part, et entre le paradigme* d'apprentissage, la perspective constructiviste de l'apprentissage et le courant de l'évaluation authentique, d'autre part.

Le paradigme d'enseignement et ses corollaires en matière d'évaluation

Le paradigme d'enseignement inspire les approches pédagogiques qui privilégient « le rôle du professeur qui enseigne à l'étudiant ». L'accent est y est mis sur la didactisation des savoirs, c'est-à-dire sur le traitement et la mise en forme qu'il faut appliquer à un contenu (une matière, une discipline) pour qu'il devienne objet d'enseignement. Cette conception a favorisé le développement de programmes d'études basés sur les objectifs pédagogiques. En convergence avec la perspective behavioriste** de l'apprentissage, on postule que l'on peut inférer correctement la qualité des apprentissages réalisés par un étudiant à partir de l'observation de ses comportements. Ce que l'étudiant doit connaître est ainsi énoncé sous forme d'habiletés spécifiques (les objectifs), hiérarchisées selon les exigences du contenu de la discipline. Dans cette optique, au moment de l'évaluation, les preuves des apprentissages d'un étudiant sont essentiellement recherchées dans la quantité d'informations qu'il aura retenues et qu'il sera capable de rappeler et dans la conformité de la restitution de ces informations avec le modèle didactique présenté

par le professeur. Ainsi, les enseignants formulent des questions ou des consignes plus ou moins élaborées et attendent les bonnes réponses ou les bons comportements, en accordant davantage de crédit au résultat plutôt qu'au processus qui a permis son élaboration. Corollairement, ils estiment que la non-atteinte d'un objectif traduit l'absence de l'apprentissage prévu. Ces différentes orientations sont très convergentes avec la perspective docimologique de l'évaluation, c'est-à-dire celle qui s'appuie prioritairement sur une théorie de la mesure. Conformément à cette option, elle privilégie comme outils de recueil des informations les procédures (tests, épreuves, examens, grilles d'observation) fournissant des informations dont l'interprétation peut être standardisée. Une procédure est réputée standardisée lorsqu'elle répond aux quatre conditions suivantes : tous les étudiants répondent aux mêmes questions ; tous les étudiants reçoivent les mêmes instructions claires et précises ; aucun étudiant ne profite d'avantages par rapport aux autres ; un système prévu de correction s'applique uniformément pour tous les étudiants. Considérées comme instruments de mesure, les procédures sont examinées et sélectionnées au regard de leurs qualités dites métrologiques : le critère de validité apprécie la capacité de l'instrument à mesurer exactement ce qu'il prétend mesurer ; le critère de fidélité apprécie la capacité de l'instrument à fournir les mêmes résultats lorsqu'il est utilisé à deux moments différents ou par deux évaluateurs différents. Selon le but poursuivi, les interprétations peuvent être de nature critériée ou normative. Fondamentalement, dans tous les cas, l'approche docimologique postule que les compétences sont décomposables en connaissances élémentaires, que l'évaluation isolée de chacune d'entre elles est pertinente et que la somme des évaluations élémentaires fournit des informations significatives pour l'évaluation du tout. Elle admet également que les connaissances sont des entités fixes et stables, qui se manifestent toujours de la même manière, indépendamment du contexte où elles s'expriment et de la compétence à l'intérieur de laquelle elles sont mobilisées³.

En résumé, l'ensemble de ces options, lorsqu'on y

* Un paradigme est un ensemble de pré-supposés scientifiques, de postulats ou de croyances, partagés à un moment donné par une communauté, qui fournit à celle-ci un cadre conceptuel pour formuler, comprendre et résoudre des problèmes.

**Se reporter à l'article de J.-L. Bernard et coll. paru précédemment dans la revue : *Pédagogie Médicale* 2001 ; 2 : 163-169

Références

adhère, conduit à développer des pratiques d'évaluation ayant en commun plusieurs points : les apprentissages des étudiants sont appréciés à partir de leurs comportements observables ; ces comportements, traduits en autant d'objectifs d'apprentissage, sont observés à l'occasion de tâches standardisées, condition essentielle de la qualité métrologique des instruments utilisés, ce qui conduit, en particulier, à séparer totalement les moments d'évaluation des moments d'enseignement et d'apprentissage ; les évaluations sont organisées à la fin de séquences autonomes, par matières ou par disciplines. L'incontestable rigueur apportée par cette approche a en contrepartie fortement contribué à morceler l'évaluation ; cet inconvénient a été fortement accentué à une époque, lorsque le courant de la pédagogie dite « par objectifs » s'est caricaturé, en faisant de la formulation des objectifs pédagogiques un but en soi.

Le paradigme d'apprentissage et ses corollaires en matière d'évaluation.

Le paradigme d'apprentissage inspire les approches pédagogiques qui privilégient « le rôle de l'étudiant qui apprend de façon individuelle, à partir des questions qu'il s'est posées, en interaction avec le professeur ». L'accent est mis sur la transformation d'informations en connaissances viables et transférables et sur l'acquisition, à partir des multiples champs qui concourent à la construction d'une compétence, d'un répertoire de connaissances grâce à des stratégies cognitives et métacognitives appropriées. Conformément aux principes issus de la psychologie cognitive* sur lesquels il s'appuie, ce point de vue considère que l'apprentissage signifiant est une construction personnelle étroitement liée à l'organisation des connaissances et que les compétences résultent de l'intégration de connaissances de faits et de connaissances d'action, au sein de schémas opératoires nécessaires à l'accomplissement de tâches complexes. En cohérence avec cette perspective, le courant de l'évaluation dite « authentique » considère fondamentalement que les moments et les contextes où se produisent les apprentissages des étudiants constituent le cadre privilégié permettant de recueillir des informations à leur sujet³. Il plaide dès lors pour qu'il n'y ait plus de clivage entre les situations d'enseignement et d'apprentissage et les situations d'évaluation.

Considérant que toute connaissance dont on vérifie l'apprentissage a vocation à être réutilisée, elle se préoccupe d'obtenir des informations qui permettent de supposer que les connaissances construites seront transférables aux contextes prévisibles de réutilisation. En rupture avec les postulats de décomposabilité et de décontextualisation évoqués précédemment, cette approche stipule au contraire qu'on ne devrait pas évaluer les connaissances en dehors des compétences attendues, ni dans des contextes qui ne seraient pas significatifs⁴. Elle admet, en effet, que la non-manifestation d'un apprentissage n'est pas systématiquement synonyme de son absence chez l'apprenant mais qu'elle peut aussi signifier que le contexte, pour des raisons liées par exemple à l'inertie et au caractère non transférable des connaissances concernées, ne permet pas sa révélation. Ce que l'étudiant doit connaître est ainsi défini à partir de tâches réelles et complexes qui sont nécessaires à l'exercice d'un rôle ou d'une fonction. Dans cette optique, au moment de l'évaluation, les preuves des apprentissages d'un étudiant seront recherchées au niveau de la qualité de l'organisation et de la transférabilité des connaissances construites et des compétences développées. La référence à des savoirs savants et à des contenus disciplinaires ne disparaît pas pour autant mais, d'une part, leur seule maîtrise n'est pas tenue pour être un indicateur suffisant de la viabilité des apprentissages et, d'autre part, les notions de pluri, d'inter et de transdisciplinarité deviennent majeures. Les instruments nécessaires à l'évaluation des connaissances et des compétences s'appuient sur des tâches dont l'authenticité (en termes de complexité, de multi-dimensionnalité, de caractère complet) est la caractéristique recherchée. Cette approche admet que la tâche d'évaluation puisse exiger une certaine collaboration avec les pairs, n'impose aucune contrainte de temps fixée arbitrairement lors de l'évaluation des compétences, prescrit au correcteur de ne tenir compte que des erreurs importantes dans l'optique de la construction des compétences et accepte que l'auto-évaluation fasse partie de l'évaluation⁹. Au critère de validité métrologique se substitue le critère de validité écologique, qui fait référence à la généralisabilité et à la transférabilité des apprentissages que ces tâches permettent d'inférer ; il s'agit de déterminer si les informations issues d'un échantillon de travaux et

* Se reporter à l'article de J.-L. Bernard et coll. (2^e partie) paru précédemment dans la revue : *Pédagogie Médicale* 2001 ; 2 : 235-241.

de productions d'un étudiant sont appropriées pour porter un jugement à l'égard des connaissances et des compétences en question, valable au-delà du contexte d'évaluation. Au critère de fidélité sont préférés les critères plus qualitatifs de transparence et de constance. En résumé, l'ensemble de ces options, lorsqu'on y adhère, conduit à développer des pratiques d'évaluation ayant en commun plusieurs points : les apprentissages des étudiants sont appréciés à partir des tâches complexes qu'ils sont capables d'exécuter ; pour être judicieusement informatives des connaissances et des compétences réellement développées, ces tâches doivent posséder un haut degré d'authenticité, ce qui conduit à enchâsser les situations d'évaluation à l'intérieur de situations d'enseignement et d'apprentissage complexes et transdisciplinaires ; les connaissances et les compétences sont appréciées de façon récurrente, à l'occasion de tâches multiples ; la détermination des critères (multiples) permettant de renseigner sur les performances attendues tient compte à la fois de la base cognitive des étudiants, de la puissance des situations d'évaluation et des compétences visées ; dans cette perspective, la distinction entre évaluation formative et évaluation sommative devient en grande partie obsolète³. L'indiscutable fécondité apportée par cette approche ne doit pas occulter les difficultés qu'elle induit. Elle exige, par exemple, l'acceptation du paradigme d'analyse qualitative dans la démarche de recherche de la « preuve », dont ne sont familiers ni les médecins, dont la formation est quasi-exclusivement inscrite la logique scientifique positiviste, ni l'administration, par nature plus facilement convaincue par les modèles technocratiques ; en raison de la nature même du processus d'évaluation suggéré, qui en fait sa richesse et sa puissance, les problèmes de fidélité inter-correcteurs, par exemple, ne sont pas à ce jour totalement résolus.

Les aspects opérationnels de l'évaluation des apprentissages

Schématiquement, la planification d'une démarche d'évaluation consiste à prendre un certain nombre de décisions concernant les raisons de l'évaluation (Pourquoi évaluer ?), l'objet de l'évaluation (Quoi évaluer ?), le moment où ces informations seront recueillies (Quand évaluer ?), les méthodes et les outils

de recueil et d'interprétation des informations (Comment évaluer ?)¹⁰. On aura compris que certaines approches pédagogiques et évaluatives tolèrent que ces différentes questions soient formulées d'une manière chronologiquement linéaire et que d'autres considèrent au contraire que ces questions et les réponses qu'elles appellent sont étroitement interdépendantes.

Pourquoi évaluer ?

Même si nous avons évoqué les limites d'une telle distinction, il s'agit, au premier degré, de décider si l'évaluation intervient dans un but formatif ou dans un but sommatif, c'est-à-dire de déterminer si la démarche et son résultat doivent servir à l'étudiant, au professeur, à l'institution, à la société ou à l'ensemble de ces interlocuteurs. Au delà de cette alternative, les motifs qui sous-tendent une opération d'évaluation sont multiples : favoriser la motivation de l'étudiant, dépister ses conceptions erronées et l'aider à les corriger, lui donner des informations sur la progression de ses apprentissages, orienter ses stratégies d'apprentissage, permettre sa promotion.

Quoi évaluer ?

Poser cette question, c'est aussi poser celle du « Quoi enseigner ? ». A cet égard, les choix faits concernant respectivement le paradigme d'enseignement ou le paradigme d'apprentissage suggèrent des approches sensiblement contrastées mais potentiellement complémentaires.

En mettant l'accent sur les savoirs codifiés, les orientations du paradigme d'enseignement prescrivent à l'évaluateur de vérifier que le contenu de la matière est adéquatement et exhaustivement couvert par les examens. Ce contenu est analysé en classant les informations qui le composent selon chacun des trois grands domaines cognitif (les « connaissances » ou encore le « savoir » selon d'autres terminologies), psychoaffectif (les « attitudes » ou encore le « savoir être ») et psychomoteur (les « habiletés » ou encore le « savoir-faire ») ; à l'intérieur de chaque domaine, les comportements observables attendus de l'étudiant sont décrits par des objectifs d'apprentissage en utilisant une hiérarchisation selon des niveaux de complexité croissante, dénommés niveaux taxonomiques. Des taxonomies très complètes ont été décrites pour chacun des trois domaines -telle

Références

celle de Bloom pour le domaine cognitif- mais, en pratique, des taxonomies simplifiées sont utilisées (Tableau 1). Les plans de construction d'un examen sont réalisés grâce à des tableaux de spécification à double entrée, qui permettent de vérifier que les instruments d'évaluation choisis (par exemple, l'ensemble des questions d'un examen) explorent bien l'ensemble des notions enseignées et l'ensemble des niveaux taxonomiques visés dans chaque domaine d'apprentissage, au prorata de leurs importances respectives dans le contenu d'enseignement tel qu'il a été planifié (validité de contenu). En réalité, les taxonomies des apprentissages ont d'abord été construites pour décrire et comprendre le développement des individus en situation d'apprentissage. Leur instrumentalisation docimologique excessive a conduit parfois à oublier que les personnes agissent comme un tout et qu'à ce point de vue, la distinction entre les différents niveaux et surtout les domaines taxonomiques est artificielle¹.

En mettant l'accent sur les compétences à développer et sur la viabilité des connaissances construites, les orientations du paradigme d'apprentissage invitent l'évaluateur à se poser les questions dans des termes différents¹¹ : quels sont les types de problème que l'on souhaite que les étudiants soient capables de résoudre ? Pour ce faire : quels concepts et quels principes doivent-ils être capables d'appliquer ? Quelles habiletés

cognitives et métacognitives doivent-ils développer ? quelles habiletés sociales et quelles attitudes doivent-ils développer ? La compétence est par nature un état cognitif dynamique interne aux individus, non directement observable ; elle ne peut être décrite qu'en s'appuyant sur une construction théorique. Dès lors, la pertinence des tâches d'évaluation résulte de leur capacité à fournir des informations documentées sur l'ensemble des connaissances qui participent à cette construction (validité de construit). Dans cette perspective, les différentes taxonomies des apprentissages devraient être davantage utilisées pour comprendre et décrire le développement professionnel des étudiants que pour satisfaire une orthodoxie docimologique.

Quand évaluer ?

Les moments où sont mises en œuvre des pratiques d'évaluation ne concernent pas seulement les aspects logistiques et calendaires de l'organisation des examens (contrôle continu ? examen final ou examens intermédiaires ? session régulière et/ou session de « rattrapage » ?). Ils conditionnent aussi largement l'objet de l'évaluation, en orientant celle-ci plutôt sur le processus ou sur le résultat de l'apprentissage et en déterminant la nature des compétences à évaluer. Les pratiques d'évaluation doivent, dès lors, être adaptées pour révéler,

**Tableau 1 :
Classifications taxonomiques simplifiées des apprentissages dans les trois domaines**

| Niveaux taxonomiques | Domaine cognitif (« Connaissances ») | Domaine psychoaffectif (« Attitudes ») | Domaine psychomoteur (« Habiletés ») |
|----------------------|--|---|--|
| Niveau I | Capacité à se souvenir de faits | Capacité à faire preuve de réceptivité à l'égard d'une autre personne | Capacité à imiter un geste |
| Niveau II | Capacité à interpréter des données | Capacité à répondre à la demande affective d'une autre personne | Capacité à contrôler un geste avec efficacité |
| Niveau III | Capacité à appliquer des connaissances à la solution d'un problème | Capacité d'intérioriser un sentiment | Capacité à accomplir un geste de manière automatique et efficiente |

selon les cas et par exemple, des compétences partielles d'étudiants de type novice en fin de cursus pré-clinique, des compétences plus achevées mais indifférenciées en fin de cursus clinique « prégradué » ou des compétences abouties et différenciées d'étudiants de type expert en fin de troisième cycle ou de cursus postgradué.

Comment évaluer ?

Le choix des instruments d'évaluation peut sembler difficile tant ceux-ci sont nombreux. Tous les procédés classiques d'examen, transposés aux spécificités de la formation médicale, sont utilisés dans les facultés de médecine. De nombreux outils ont aussi été plus ou moins codifiés pour l'évaluation des apprentissages au cours des stages hospitaliers. Le tableau 2 (a et b) décrit brièvement les procédés les plus utilisés en éducation médicale. L'analyse critique des forces et des limites de chacun dépasserait les limites imparties à cet article. L'évaluation de compétences finales, dans la perspective de la délivrance d'une licence d'exercice professionnel ou d'une certification, pose notamment de nombreux problèmes qui sont loin d'être résolus^{12, 13} et d'autres contributions à paraître dans la revue les aborderont. D'une façon générale en évaluation sommative, la détermination des standards, c'est-à-dire des seuils de réussite au-delà desquels les étudiants sont admis, reste une démarche arbitraire¹⁹ ; certains outils statistiques sont utilisés mais la décision finale résulte toujours d'un accord professionnel entre experts du domaine. Les différents procédés de recueil d'informations ne tirent pas tant leur légitimité de leurs caractéristiques techniques intrinsèques que de la cohérence de leur choix, de leur mise en œuvre et de l'interprétation de leurs résultats avec les orientations pédagogiques et les intentions du programme de formation.

Il est, par ailleurs, tout à fait essentiel que les procédures d'évaluation sélectionnées soient compatibles avec les contraintes pragmatiques de l'établissement et de l'institution au sein desquels le programme de formation est mis en œuvre. La planification et la mise en œuvre de procédures d'évaluation des apprentissages mobilisent de multiples ressources humaines, en enseignants mais aussi en personnel de soutien, qui doivent être disponibles. Elles ont aussi un coût financier qui peut parfois être réhibitoire. Conformément aux exigences de cohérence pédagogique, elles peuvent nécessiter également une formation pédagogique approfondie des enseignants. Enfin, quels que soient les paradigmes auxquels

elles se réfèrent, ces procédures doivent être assorties de garanties de rigueur qui confèrent au jugement évaluatif porté un caractère exact et juste. Il en va de la validité déontologique de la démarche évaluative ; celle-ci est fortement influencée par le contexte académique, social et culturel concerné, qui peut conditionner fortement l'acceptabilité de certaines pratiques. A cet égard, il est pertinent d'admettre, par exemple, que les étudiants français nourrissent toujours une forte suspicion à l'égard des examens oraux, dont ils dénoncent le risque de dérive népotique ; il est loyal de reconnaître aussi que le recours aux épreuves utilisant des patients standardisés a été favorisé par le dispositif nord-américain de soins conventionnés (*managed care*) et par l'influence que les puissantes compagnies d'assurance ont développée, dans ce contexte, à l'égard des facultés de médecine et de la formation médicale. Dans aucun pays en tout cas, les enseignants n'échapperont à l'exigence sociale, de plus en plus explicitement formulée, d'être en mesure d'argumenter, de façon transparente, au nom de quoi et comment ils prennent leurs décisions d'évaluation.

Synthèse et perspectives

Alors même que l'évaluation est centrale dans le processus d'enseignement et d'apprentissage, le soin avec lequel les démarches d'évaluation sont développées n'est pas toujours parfaitement documenté. Il n'est pas rare de constater que, dans un programme de formation donné et au fil du temps, les pratiques évaluatives répondent majoritairement à des routines, largement déconnectées de leur raison d'être et qu'elles ne résultent pas -ou plus- d'un choix explicite, décidé à l'issue d'un raisonnement pédagogique.

Sans rentrer dans les aspects technologiques de l'évaluation, qui n'étaient pas l'objet de cette mise au point, quelques recommandations générales peuvent être dégagées concernant les systèmes d'évaluation :

- Développer un dispositif d'évaluation en ayant à l'esprit les conséquences prévisibles de ces procédures sur les apprentissages des étudiants

Il est toujours préférable de privilégier les dispositifs qui favorisent les apprentissages en profondeur. Les procédés qui permettent d'avoir accès à l'organisation des connaissances et aux liens conceptuels que les étudiants ont construits, ceux qui font appel à la réutilisation pertinente de connaissances plutôt qu'à leur réci-

Tableau 2 :
Panorama des principaux instruments d'évaluation des apprentissages
utilisés en formation médicale initiale

a- Tâches académiques

- Question rédactionnelle¹

Question ouverte nécessitant de la part de l'étudiant un développement personnel plus ou moins réflexif. La correction peut en être assurée avec des grilles plus ou moins explicites, plus ou moins ouvertes ou fermées.

- Question à réponse ouverte et courte^{1*} (QROC)

Question ouverte nécessitant de la part de l'étudiant une réponse courte et précise, laquelle constitue le barème exact de correction.

- Question à choix multiples^{1*} (QCM)

Question dont la partie initiale (le tronc) peut prendre la forme d'une question directe ou d'un énoncé incomplet et qui comprend un certain nombre de réponses suggérées parmi lesquelles l'étudiant doit choisir. La correction est binaire (vraie ou fausse).

- Schéma cognitif^{12*} (Concept mapping)

Représentations visuelles et propositionnelles de concepts thématiques, liés entre eux par une relation explicitement nommée (causalité, conditionnalité, identité,...) organisés hiérarchiquement, à l'égard desquelles un jugement qualitatif ou quantitatif peut être porté.

**Ces instruments d'évaluation peuvent ou non être contextualisés à des problèmes de santé. QROC et QCM le sont souvent peu.*

b- Tâches professionnelles authentiques, simulées ou modélisées

- Analyse de cas cliniques^{13, 14} [Patients Management Problem (PMP) ; Sequential Management Problem (SMP) ; Modified Essay Question (MEQ)]

Dossier clinique (sous forme papier ou sous forme numérique) dont les informations sont fournies de manière séquentielle, linéaire ou bien algorithmique, à la demande de l'étudiant, après chacune de ses réponses (ouvertes ou fermées). La démarche employée par l'étudiant est comparée à celle d'un expert ou d'un groupe de référence et l'appréciation est formulée à l'aide de scores plus ou moins combinés.

- Test de concordance de script¹⁵

Un problème clinique est soumis à l'étudiant qui doit interpréter des données et prendre des décisions. Des hypothèses d'interprétation ou de décisions lui sont fournies en même temps que des informations nouvelles lui sont communiquées. L'étudiant doit indiquer à l'aide d'une échelle l'importance du crédit ou du discredit que l'information nouvelle apporte à l'hypothèse.

- **Observation directe de l'étudiant dans une tâche professionnelle avec des tâches et/ou des patients standardisés : Examen Clinique Objectif Structuré (ECOS)¹⁶**

Une personne (bien portant, patient stabilisé, comédien) est entraînée à jouer de manière reproductible le rôle standardisé d'un patient. Les activités d'interrogatoire, d'examen physique, d'éducation et de conseil de l'étudiant sont appréciées à l'aide de « check-lists », d'échelles de cotation, élémentaires ou globales. Des tâches procédurales complémentaires exploitant divers matériels (mannequins, imagerie, résultats de laboratoire, ECG,...) sont prescrites et cotées. L'ensemble fait l'objet de « stations » d'évaluation (indépendantes ou complémentaires) codifiées.

- **Observation directe de l'étudiant dans une tâche professionnelle avec des patients non standardisés mais des grilles d'observation standardisées**

• **Formulaire d'observations longitudinales¹³**

Des échelles de notation (de 1 à 9) sont développées pour mesurer 10 dimensions : le jugement clinique, les connaissances médicales, l'interrogatoire, l'examen physique, les habiletés procédurales, les habiletés interpersonnelles, les soins médicaux, l'humanisme, le professionnalisme, la compétence générale. Des descripteurs comportementaux sont établis pour chaque seuil : performances non satisfaisantes (1-3), satisfaisantes (4-6), supérieures (7-9).

• **Clinical Evaluation Exercise¹³ (CEX)**

L'étudiant est observé pendant qu'il interroge et examine un patient puis il présente le cas et en discute la gestion. La cotation est faite à l'aide d'une échelle d'évaluation à 9 points

• **Objective Structured Long Examination Record¹⁷ (OSLER)**

L'étudiant est observé par 2 examinateurs pendant qu'il interroge et examine un patient puis il argumente sa prise en charge. Les examinateurs fondent leur jugement en appréciant 10 items, dont 4 concernent l'interrogatoire, 3 l'examen physique, 1 la démarche d'investigation, 1 la gestion et le jugement clinique. Chaque item fait l'objet d'une cotation (en 3 grades) et d'une note (en 9 paliers de 35 à 80), étayée par des descripteurs critériés. Les problèmes sont classés en trois niveaux de difficultés.

- **Outils visant à documenter des apprentissages effectués à partir de pratiques réelles :**

• **Portfolio¹⁸**

Collections organisées de matériel permettant d'obtenir des informations sur les apprentissages effectués et sur les compétences développées par un étudiant au cours d'une période plus ou moins longue d'apprentissage. Elles contiennent : - une compilation de scripts de problèmes complets ou de certains événements (échecs, incidents,...), - des notes d'analyse bibliographique critique, des résumés d'entretiens avec des personnes ressources ou de discussions collégiales, des descriptions de projets en cours, des « journaux de bord » d'activités professionnelles, des vidéoclips d'entretiens avec des patients ou de tâches procédurales,... ; - une analyse réflexive développée à partir de cet échantillon de problèmes et de cette banque documentaire

• **Audit de dossiers médicaux¹³**

Des « biopsies » sont effectuées parmi les dossiers des patients pris en charge par les étudiants. Les informations « objectives » du dossier (observations, pancartes, courrier, prescriptions, feuilles de surveillance...) sont analysées et appréciées.

Références

tation sur commande, ceux qui se donnent les moyens d'avoir accès au processus d'un raisonnement et non seulement à son résultat sont cohérents avec cette intention. D'une façon générale, les situations d'évaluation doivent toujours comporter un niveau minimum de complexité cognitive.

- Délimiter ce qui doit être évalué en raison de sa pertinence et de son importance par rapport aux compétences visées plutôt qu'en raison de sa facilité à être mesuré

Les concepts de validité de contenu et de fidélité sont utiles pour rappeler aux enseignants la nécessité d'une rigueur dans leur démarche évaluative. Une certaine « dictature » docimologique risque cependant de faire perdre de vue aux enseignants la nécessité que les savoirs évalués possèdent avant tout une forte pertinence et une forte légitimité par rapport à l'expertise professionnelle visée. Pour ce faire, la validation des objets de l'évaluation grâce à un processus de concertation interprofessionnelle pluridisciplinaire est souvent une approche féconde.

- Sélectionner des modalités d'évaluation qui renseignent valablement sur le niveau de performance réellement attendu des étudiants

Le dispositif de formation médicale initiale assiste l'étudiant dans son cheminement du stade de novice vers celui de professionnel expert. Certaines procédures de recueil d'informations, telles que les observations directes standardisées utilisant des grilles de vérification systématique (*check-lists*) sont adaptées pour évaluer des comportements novices ; elles peuvent en revanche se révéler impropres à apprécier des comportements experts. Cet inconvénient peut être pallié de diverses manières (par exemple, en complétant l'observation directe par un protocole écrit ou oral de verbalisation du processus). L'évaluation de compétences approchant le niveau d'expertise doit reposer sur des situations d'évaluations ayant une réelle complexité cognitive et professionnelle.

- Exposer les étudiants à un nombre et à une variété suffisants de situations d'évaluation

Lorsque les situations d'évaluation sont enchâssées au

sein des situations d'enseignement et d'apprentissage, elles sont par nature récurrentes. Dans les autres cas, il importe de prévoir un nombre suffisant de situations d'évaluation, à condition de s'assurer qu'elles explorent bien des connaissances et des compétences complémentaires. Ceci répond au souci que les jugements évaluatifs portés soient généralisables et vise à pallier les limites tenant à la spécificité de contenu de très nombreuses situations d'évaluation (leurs résultats sont peu transférables).

- Développer des dispositifs d'évaluation utilisant des méthodes et des instruments que l'on maîtrise, dont on connaît les forces et les limites, compatibles avec les allocations en ressources du programme de formation (coûts humains et financiers) et acceptés par les acteurs concernés (les étudiants et les enseignants)

Les meilleurs outils peuvent perdre tout ou partie de leurs qualités intrinsèques lorsqu'ils sont utilisés de façon inadéquate, par exemple avec des ressources humaines insuffisantes. Il est inutile de compliquer inutilement les dispositifs d'évaluation. Ainsi, les barèmes développés pour l'attribution des scores et des notes doivent rester simples ; les barèmes compliqués, qui tentent d'introduire une pondération très différenciée des différentes composantes de l'apprentissage, fournissent le plus souvent des résultats redondants¹⁴. Quels que soient les dispositifs mis en place, le jugement des enseignants doit être, *in fine*, fortement sollicité et ne pas être prisonnier des instruments utilisés : ce n'est pas tant la mesure qui doit être valide et fidèle que le jugement qui est fait à partir d'elle.

- Rendre explicite aux étudiants, dès le début de la séquence d'enseignement et d'apprentissage, la nature des tâches d'évaluation et les critères qui seront pris en compte pour la correction et/ou pour la détermination de la réussite

Il n'est pas acceptable que les étudiants soient placés en situation divinatoire par rapport aux intentions des enseignants. La formulation d'objectifs d'apprentissage explicites répond à cette exigence. Un formalisme excessif a parfois entaché les règles canoniques de leur rédaction ; il est, sans doute, plus essentiel que ces

objectifs soient présentés en perspective avec les compétences visées et il est par ailleurs tout à fait nécessaire d'en vérifier la compréhension par les étudiants.

Les ancrages conceptuels qui ont été rappelés concernant tant les paradigmes d'enseignement et d'apprentissage que les courants théoriques de l'évaluation ne doivent pas conduire à développer une vision manichéenne et réductrice du champ de l'évaluation et de ses pratiques. Dénoncer les excès de certains positionnements ou de certaines pratiques ne veut pas dire qu'il faut nécessairement « jeter le bébé avec l'eau du bain ». D'une part, les théories et les modèles psychologiques qui inspirent les pratiques pédagogiques, ainsi que les courants sociaux et culturels au sein desquels ces pratiques prennent place, se développent à la fois en rupture et en continuité les uns par rapport aux autres. Par ailleurs, même si le paradigme d'apprentissage est celui qui inspire aujourd'hui les changements pédagogiques considérés comme les plus innovants en éducation médicale -et dont « l'apprentissage par problèmes » est devenu depuis une trentaine d'années le dispositif le plus emblématique-, force est d'admettre que, mêmes dans les facultés de médecine qui se sont engagées dans une telle mutation, de nombreux formats de planification et d'intervention pédagogiques restent profondément ancrés dans le paradigme d'enseignement. On pourrait ajouter que si le développement de l'expertise médicale nécessite effectivement des apprentissages complexes, certaines habiletés ou aptitudes tout à fait nécessaires à la pratique médicale courante peuvent résulter d'apprentissages moins exigeants et que, pour ceux-là, il ne serait peut-être ni judicieux, ni économique de bouleverser les dispositifs de formation et d'évaluation. Il faut, en outre, souligner les problèmes de l'interrelation « recherche-pratique-formation » et constater l'existence d'un décalage temporel important entre le développement des modèles théoriques issus de la recherche et leur traduction dans les pratiques²⁰ : ainsi, la psychologie cognitive a donné lieu à des applications fort cohérentes et fort fécondes en termes de pratiques d'enseignement et d'apprentissage mais il faut admettre que sa transposition opérationnelle en matière de pratiques évaluatives est aujourd'hui en retard. Plusieurs problèmes méthodologiques devront être résolus, par exemple, pour que des pratiques issues du courant de l'évaluation authentique s'imposent, à

une large échelle, au sein des procédures à visée certificative. Les facultés de médecine ne peuvent ainsi pas s'affranchir de façon radicale et sans inconvénient des procédures développées ou prescrites par les collèges professionnels ou par les gouvernements.

Conclusion

La démarche évaluative est fondamentalement une pratique sociale, développée par des enseignants qui agissent en tant que professionnels mandatés. De ce fait, elle n'est jamais réductible à l'application de recettes stéréotypées. Conçue comme un processus de résolution de problème complexe, elle partage avec la démarche clinique la double exigence de rigueur et de pertinence ainsi que la nécessité de prendre systématiquement en compte des dimensions éthiques pour, parfois, affronter des conflits de valeurs. Elle nécessite que les enseignants disposent des outils conceptuels leur permettant de donner un sens aux actes d'évaluation qu'ils posent, en anticipant notamment les conséquences que ces actes auront sur les apprentissages des étudiants. Pour ce faire, elle implique nécessairement un positionnement théorique, qui n'a rien de définitif, qui possède ses forces et ses limites, mais qui est analysé, à un moment donné et dans un contexte éducatif déterminé, comme étant le compromis pédagogique le plus cohérent. Dans cette perspective, la réflexion sur les instruments doit être éclectique et ne jamais confondre les buts et les moyens : « Un ouvrier qui possède pour seul outil un marteau finit toujours par considérer que tous ses problèmes sont des clous »²¹.

Références

Références

1. Legendre R. *Dictionnaire actuel de l'éducation*, Montréal-Paris : Guérin-Eska 1993.
2. Scallon G. *Evaluation formative et psychologie cognitive : mouvances et tendances*. In : Grégoire L. (Ed.). *Évaluer les apprentissages. Les apports de la psychologie cognitive*. Paris, Bruxelles : De Boeck Université 1996 : 159-173.
3. Tardif J. *Une évaluation authentique*. In : Tardif J. *Intégrer les nouvelles technologies de l'information. Quel cadre pédagogique ?* Paris : ESF Editeur 1998 : 80-106.
4. Tardif J. *L'évaluation dans le paradigme constructiviste*. In : Hivon R. (Ed.). *L'évaluation des apprentissages*. Sherbrooke (QC) : Editions du CRP : 27-56.
5. Louis R. *Les différentes approches de l'évaluation des apprentissages*. In : Louis R. *L'évaluation des apprentissages en classe. Théorie et pratique*. Laval (QC) : Editions Etudes Vivantes 1999 : 37-49.
6. Tardif J. *Motivation scolaire*. In : Tardif J. *Pour un enseignement stratégique. L'apport de la psychologie cognitive*. Montréal (QC) : Editions Logiques 1992 : 87-152.
7. Viau R. *La motivation en contexte scolaire*. Paris, Bruxelles : De Boeck Université 1996.
8. Wiggins G. *Curricular coherence and assessment : making sure that the effects matches the intent*. In : Beane JA. (Ed.). *Toward a coherent curriculum. The ASCD Yearbook*. Alexandria (VA) : Association for supervision and curriculum development 1995 : 101-119.
9. Wiggins G. *Standards, Not standardization : Evoking quality of student work*. *Educational Leadership* 1991 ; 48 : 18-25.
10. Hivon R, Louis R. *L'évaluation des compétences*. In Hivon R. et Tardif J. (Eds.). *Le développement de l'expertise professionnelle. Ecole internationale d'été en pédagogie universitaire. Cahier Pédagogique*. Sherbrooke : Université de Sherbrooke et Presses Inter Universitaires 1995 : 68-85.
11. Huba ME, Freed JE. *Assessing student's ability to think critically and solve problems*. In : Huba ME, Freed JE (Eds.). *Learner-centered assessment on college campuses. Shifting the focus from teaching to learning*. Needham Heights (MA) : Allyn and Bacon, 2000 : 201-232.
12. Pinto AJ, Zeitz H. *Concept mapping : a strategy for promoting meaningful learning in medical education*. *Med Teacher* 1997 ; 19 : 114-121.
13. Holmboe ES, Hawkins RE. *Methods for evaluating clinical competence of residents in internal medicine : a review*. *Ann Intern Med* 1998 ; 129 : 42-48.
14. Newble D, Norman G, Van der Vleuten. *Assessing clinical reasoning*. In : *Clinical reasoning in the Health Professions*. J Higgs and M. Jones (Eds), Butterworth Heineman, Oxford : 2000
15. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. *The Script Concordance test : a tool to assess the reflective clinician*. *Teach Learn Med* 2000 ; 12 : 189-195.
16. Harden RM, Gleeson FA. *Assessment of medical competence using an Objective Structured Clinical Examination*. *Med Educ* 1979 ; 13 h 39-54.
17. Gleeson F. *Assessment of clinical competence using the Objective Structured Long Examination Record*. *Med Teach* 1997 ; 19 h 7-14.
18. Challis M. *Portfolio-based learning and assessment in medical education*. *Med Teach* 1999 ; 21 : 370-386
19. Friedman Ben-David M. *AMEE Guide n° 18 : Standard setting in student assessment*. *Med Teach* 2000 ; 22 : 120-130
20. Bligh J. *Assessment : the gap between theory and practice*. *Medical Education*, 2001 ; 35 : 312.
21. Anonyme, cité par M. Samuelson. *Communication personnelle*.