



Section des unités de recherche

Évaluation des unités de recherche par l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES) : vers une amélioration de la notation¹

— Décembre 2011 —

¹ Version adoptée, à l'unanimité des membres présents du conseil de l'AERES moins une abstention, lors de la séance du 08/12/2011.

La loi confie à L'AERES (art. L. 114-3-1-2°) une mission d'évaluation des activités de recherche conduites par les unités de recherche des établissements et organismes de recherche, et des établissements d'enseignement supérieur et de recherche. La méthode d'évaluation retenue par l'AERES s'appuie sur un travail d'auto-évaluation réalisé par l'entité évaluée, puis sur une évaluation externe, indépendante, collégiale et transparente, effectuée par des pairs. Il en résulte un rapport textuel public² auquel sont adjointes les observations de l'unité en retour à ce rapport et un ensemble de notes dont la forme a évolué au cours des campagnes successives.

Alors que l'AERES termine, en 2011, un premier cycle d'évaluation des près de 3000 unités de recherche identifiées en France, il y avait lieu de s'interroger sur l'avenir de la notation des unités de recherche par l'AERES.

Ce document fait le point sur la politique de notation de l'AERES dans ce domaine (p. 2). Se basant sur l'analyse des retours d'expérience (p. 3 et 4) et la comparaison avec les pratiques dans d'autres pays européens (p. 7), il redéfinit les objectifs de l'évaluation et, plus spécifiquement, de la notation, en relation avec les utilisateurs de celles-ci (p. 5 et 6). Sur la base de ces éléments, il propose des pistes d'amélioration pour le système actuel de notation, en particulier en ce qui concerne la note globale (p. 8).

1 - L'évolution de la notation dans le cadre de l'évaluation des unités de recherche par l'AERES

Dès la mise en place de l'AERES en 2007, l'évaluation des unités de recherche a comporté une notation. Au fur et à mesure des vagues d'évaluation et des retours d'expérience organisés, la notation accompagnant l'évaluation a évolué (Tableau 1). Ainsi, à la suite de l'avis favorable du Conseil de l'AERES, le 09 octobre 2008, la notation multicritères a été complétée par une notation globale de l'unité évaluée.

Tableau 1 : Évolution de la notation des unités de recherche dans le cadre des évaluations conduites par l'AERES

	Vague C (2007-2008)	Vague D (2008-2009)	Vague A (2009-2010)
Mode de notation	Globale sur la base d'une grille remplie par les comités d'experts	Globale + multicritères	Globale + multicritères, avec grille et échelle explicitées La notation est faite après réception des observations sur le rapport textuel en retour de l'unité.

Actuellement, la notation, en tant qu'élément du processus d'évaluation des unités de recherche, porte sur l'unité de recherche et, le plus souvent, sur les équipes qui la composent. Elle se matérialise par :

- un ensemble de quatre notes (notation multicritères), portant respectivement sur : qualité scientifique et production ; rayonnement et attractivité, intégration dans l'environnement ; gouvernance et vie du laboratoire ; stratégie et projet scientifique ;
- une note globale, distincte des notes associées à chacun des critères : elle traduit la synthèse de l'opinion collégiale des présidents de comité d'experts du groupe des unités qui, dans un domaine de recherche donné, ont été évaluées au cours de la campagne en cours.

Ces notes sont déclinées sur quatre niveaux A+, A, B et C, qu'il s'agisse des notes multicritères ou de la note globale, sans proportion préétablie pour chacun des quatre niveaux.

Les rapports d'évaluation produits par l'AERES, y compris les contributions des experts, les observations des « évalués » et les notes, sont rendus publics par l'AERES.

² Accessible sur le site internet de l'AERES.

2 - Des insatisfactions ont été exprimées vis-à-vis de la notation des unités de recherche

Dans l'ensemble, le principe de la notation est accepté par les parties prenantes (notamment par les entités évaluées et les décideurs/financeurs au niveau des établissements/organismes/ministères /collectivités territoriales). L'AERES a toutefois recensé, au fil du temps, plusieurs expressions d'insatisfaction vis-à-vis de la notation, n'émanant pas seulement d'unités pour lesquelles la note a été basse.

Ces expressions peuvent résulter d'une incompréhension d'ordre méthodologique, le processus d'évaluation par les pairs méritant d'être plus largement explicité et différencié de démarches d'évaluation strictement quantitatives.

A cet égard, il convient de préciser que la phase de l'évaluation des unités de recherche de l'AERES discutée ici comprend deux étapes :

- Une étape d'analyse par les pairs (le comité des experts qui visitent l'unité), non réductible à l'utilisation de critères quantitatifs, même si les experts peuvent en faire un usage partiel. Le résultat du travail des pairs inclut alors une série d'arbitrages internes (prise en compte de connaissances contextuelles échangées dans le groupe, recalages divers pour introduire des différences d'opinion qui ne sont pas communiquées à l'extérieur, dynamique de groupe...). Cette approche est d'une grande force si la méthodologie est rigoureuse, mais le résultat reste un jugement qualitatif, délicat à retraiter après coup.
- Une étape d'appréciation comparative, lors de laquelle des notes (associées à chacun des critères et globale) sont attribuées par des panels de pairs (les présidents des comités ayant conduit l'étape d'analyse). Cette phase pourrait permettre de faire des comparaisons entre entités « comparables » (même domaine, mêmes missions, etc.). Cette étape d'appréciation condense les résultats de l'étape précédente. Elle fait appel à une approche différente. Elle mène en particulier à une réduction de l'information, ce qui en limite fortement les possibilités d'usage.

Les expressions d'insatisfaction portent, quant à elles, principalement sur :

- la prise en compte insuffisante, dans l'évaluation, de certaines activités de recherche finalisée³, ainsi que d'activités connexes à la recherche (notamment, de l'expertise venant en appui des politiques publiques). Il en résulte une difficulté à reconnaître et valoriser pleinement, dans le référentiel multicritères actuel, la grande diversité des unités dans leurs missions et leurs productions ;
- le nivellement des notes des équipes du fait de l'existence d'une note globale pour l'unité ;
- l'élaboration et l'usage de la note globale :
 - sa relation avec les notes associées à chacun des critères n'apparaît pas claire ;
 - ses modalités d'élaboration dépendent du fonctionnement de chaque collège de présidents (panel de notation). Elles souffrent d'un défaut : le jugement qualitatif émis par les pairs, lors d'un comité de visite, est difficilement transmissible après coup au panel de notation, dont la composition est par ailleurs différente de celle des comités de visite ayant procédé à l'évaluation.
 - en dépit de son caractère réducteur, la note globale semble parfois être le seul élément pris en compte lors des décisions concernant l'unité évaluée (notamment, au détriment de la prise en compte des notes associées à chacun des critères et de la lecture du rapport d'évaluation) ;
- le caractère insuffisamment distingué de l'excellence (en bref, trop de notes A+) ;
- la persistance, durant plusieurs années, des notes sur le site de l'AERES (« tatouage »), alors que la situation de l'unité évaluée a pu beaucoup changer entre temps.

Ces expressions reflètent sans doute aussi la multiplicité des destinataires des résultats de l'évaluation et la diversité de leurs attentes et usages. Elles suggèrent que la notation ne trouve sens que pour un usage spécifié. Si chacun s'accorde sur le fait que l'évaluation doit avoir pour effet l'amélioration, il apparaît aussi que le signal ou l'information résumée, que constituent les notes associées à chacun des critères, joue bien son rôle vis-à-vis de nombreux destinataires.

Certaines de ces expressions invitent enfin à ce que le système actuel de notation progresse en termes de robustesse et de reproductibilité. La robustesse peut être améliorée, si les notes ne sont pas trop sensibles à la prise en compte d'éléments marginaux ou à de petites variations des critères. L'amélioration de la reproductibilité nécessiterait que les notes soient indépendantes de la composition des panels de présidents de comité de visite, qui les attribuent. Cette progression sera aussi rendue possible, si les critères et les échelles de notation sont définis avec suffisamment de précision.

³ Selon le domaine, on parlera de recherche finalisée, appliquée, translationnelle, technologique, ou clinique...

3 - En vue d'amélioration, un chantier a été ouvert à l'AERES sur la notation des unités de recherche

En juin 2011, un chantier a donc été ouvert, qui s'est concrétisé, à ce stade, par :

- une première réunion de travail interne à l'AERES, le 22 juin 2011 ;
- une première séance de travail au sein du conseil de l'AERES, le 7 juillet 2011 ;
- deux séminaires de travail interne à l'AERES, avec participation de membres du conseil de l'AERES, les 8 et 21 septembre 2011 ;
- une deuxième séance de travail au sein du conseil de l'AERES, le 22 septembre 2011 ;
- une troisième séance de travail au sein du conseil de l'AERES, le 13 octobre 2011 ;
- une quatrième séance de travail au sein du conseil de l'AERES le 10 novembre 2011 ;
- une cinquième séance de travail au sein du conseil de l'AERES le 08 décembre 2011.

Lors de cette dernière séance de travail au sein du conseil de l'AERES, le présent document a été soumis à approbation par le conseil de l'AERES, afin de présenter différents modes d'amélioration de la notation, dans la perspective de leur mise en place lors du cycle à venir d'évaluation des unités de recherche.

Il a été estimé qu'à un nouveau cycle d'évaluation de cinq ans devait correspondre un dispositif homogène d'évaluation et de notation. Le nouveau dispositif retenu (voir en particulier la proposition figurant au point 7.3) devra être opérationnel avant le 15 février 2012.

4 - Vers une meilleure identification des objectifs de l'évaluation des unités de recherche visés par l'AERES

Ces évaluations se veulent constructives. Les évaluations conduites par l'AERES sont destinées en priorité à **permettre à l'unité évaluée d'identifier des pistes d'amélioration**. Le rapport d'évaluation est alors destiné à servir de point d'appui en vue de décisions à prendre par l'unité évaluée en accord avec ses tutelles.

Un deuxième objectif est que les évaluations des unités de recherche, conduites par l'AERES, **facilitent les décisions susceptibles d'être prises par des acteurs extérieurs à l'unité évaluée**. Cette facilitation doit se traduire par une présentation détaillée des résultats de l'évaluation, en tenant compte des différents critères susceptibles d'intéresser le décideur, compte tenu de sa stratégie d'ensemble, des moyens dont il dispose et des éléments de contexte de sa décision. Ces décisions peuvent toucher à l'existence même de l'unité, mais aussi à sa direction ou à son financement. En particulier, la Loi prévoit que « l'Etat tient compte des résultats de l'évaluation réalisée par l'AERES,..., pour déterminer les engagements financiers qu'il prend envers les établissements.... » (art. L. 311-2 du code de la recherche et art. L. 711-1 du code de l'éducation).

Un troisième objectif est que ces **évaluations contribuent à l'information de tous ceux, qui n'ont pas de rôle de décision vis-à-vis de l'unité évaluée**, mais qui sont intéressés à connaître les résultats de l'évaluation. Ceux-ci doivent donc leur être présentés de la manière la plus simple et la plus aisément compréhensible.

5 - Un recensement plus précis des destinataires des évaluations des unités de recherche organisées par l'AERES

Compte tenu des objectifs de l'évaluation identifiés au chapitre précédent, il est possible de dessiner l'ensemble des destinataires potentiels de l'évaluation des unités de recherche sous la forme de trois cercles concentriques.

Le premier cercle est constitué par **les responsables et l'ensemble des personnels de l'unité évaluée**, mais aussi par les experts ayant participé à l'évaluation. Tous ces destinataires sont touchés directement, lors de la réception du rapport d'évaluation, par le rapport lui-même, les notes, les contributions éventuelles des experts, les observations de l'unité évaluée, et par la publicité faite à l'ensemble du rapport final. C'est la raison pour laquelle l'AERES est attentive à la manière selon laquelle sont rédigés les différents documents, qui doit être empreinte de respect.

Le second cercle des destinataires est constitué **des personnes susceptibles de jouer un rôle de décideur**, public ou privé, vis-à-vis de l'unité évaluée, qu'il s'agisse de son existence, de sa direction, de ses moyens, notamment financiers ou, indirectement, de la mise en place, par exemple, d'un partenariat. Devant parfois prendre de nombreuses décisions dans un bref délai, ces destinataires peuvent souhaiter une information présentée de manière à leur permettre de prendre rapidement la décision la plus adaptée à la stratégie qu'ils ont à mettre en œuvre.

Le troisième cercle est fait de destinataires très nombreux et divers. Si certains peuvent souhaiter des informations détaillées, d'autres, en tant que simples citoyens ou parce qu'ils exercent des responsabilités publiques, veulent se faire une idée sur l'état de la recherche dans une région ou la qualité d'une unité donnée. Les résultats doivent alors leur être présentés de la manière la plus synthétique possible. Une telle présentation n'est pas intrinsèque à l'évaluation, mais sa production peut être un vecteur de confiance et répondre à une demande sociétale importante.

6 - La prise en compte des modalités de notation utilisées dans d'autres pays

Dans la mesure où le développement de l'espace européen de la recherche fait partie des projets soutenus dans le cadre de l'Union européenne, il est nécessaire, s'agissant de l'évaluation des unités de recherche en France, de tenir compte, autant que possible, des modalités d'évaluation mises en place au sein d'autres pays européens.

Cette prise en compte est importante, afin de :

- permettre un regard plus homogène sur la qualité de la recherche en Europe ;
- faciliter l'analyse des partenariats existant entre unités de recherche.

L'évaluation de la recherche en Europe est le plus souvent effectuée à l'échelle des pays. A un grain plus fin, ce sont souvent les universités, soit dans leur ensemble, soit par département, qui sont évaluées. C'est le cas en particulier en Allemagne (évaluation, par exemple, par le *Centrum für Hochschule Entwicklung* (CHE)⁴) ou en Angleterre (évaluation par le *Higher Education Funding Council of England* (HEFCE)⁵). Aux Pays-Bas, les instituts sont évalués qualitativement et donc ne sont pas notés, tandis que les groupes de recherche le sont⁶. Le nombre des critères et de niveaux de notation varie de 3 à 5. En Espagne, les unités et les équipes de recherche sont évaluées par l'Agencia Nacional Evaluacion y Prospectiva (ANEP)⁷.

Le système européen U-Multirank⁸, destiné à l'évaluation des universités, fait référence, au premier niveau, à 15 critères regroupés en 5 rubriques. Les outils proposés par le CHE et U-Multirank sont paramétrables, c'est à dire qu'ils permettent une sélection par l'utilisateur des critères qu'il trouve les plus pertinents pour son usage.

Une présentation plus détaillée de ce qui se fait dans différents pays européens figure en Annexe 1.

7 - Propositions d'amélioration du dispositif de notation des unités de recherche

L'avis rendu sur ces propositions, a des enjeux importants :

- la responsabilité assumée de l'évaluation : la notation, en particulier, engage la responsabilité de l'évaluateur, en raison de son impact sur l'unité évaluée ;
- l'acceptabilité par les parties prenantes et l'adéquation à leurs besoins ;
- la lisibilité d'ensemble du dispositif d'évaluation et de notation ;
- la capacité d'adaptation et de prise en compte des résultats de l'apprentissage par l'AERES ;
- la prise en compte des réflexions méthodologiques conduites dans d'autres pays européens ;
- la préservation de la mémoire de l'évaluation, mais aussi la réactivité, en cas de changement net de la situation d'une unité antérieurement évaluée.

Proposition de maintien des notes associées aux différents critères avec révision des critères d'évaluation

La première principale proposition est de maintenir les notes associées aux différents critères, mais aussi de réviser, pour les améliorer, les critères d'évaluation.

Le maintien, sans changement, du mode de notation multicritères permettra aussi d'éviter la confusion qui, au regard des évaluations des années passées, s'instaurerait en cas de changement du nombre de niveaux de notation ou de la dénomination des niveaux des notes multicritères.

Il est, en revanche, nécessaire de réviser les critères d'évaluation. Il s'agit, en particulier, que cette révision permette de mieux tenir compte de la diversité des missions et productions de la recherche. Dans ce cadre, méritent d'être examinés, par exemple : (i) le rapport du groupe EREFIN s'agissant de la recherche finalisée ; (ii) les travaux conduits en Grande-Bretagne sur le thème de l'impact de la recherche dans le cadre du passage du *Research Assessment Exercise* de 2008 au *Research Evaluation Framework (REF)* pour 2014 ; (iii) les travaux engagés au sein de l'AERES sur l'évaluation de la recherche clinique, etc.

La révision des critères actuellement utilisés par l'AERES, sera rapidement déclenchée. Elle s'appuiera, notamment, sur les différents travaux produits par l'Agence et sur les contributions qui lui ont été envoyées au cours des dernières années (document EREFIN, rapport R. Chabbal, document de travail transmis par M. Lefort en septembre 2011, ...). Elle sera, dans tous les cas, opérationnelle pour la vague D.

⁴ Source CHE : <http://www.daad.de/deutschland/hochschulen/hochschulranking/06543.en.html>

⁵ Source HEFCE : <http://www.hefce.ac.uk/research/ref/>

⁶ Source *Standard Evaluation Protocol* (The Netherlands) : http://www.knaw.nl/content/Internet_KNAW/publicaties/pdf/20091052.pdf

⁷ Source ANEP : <http://www.micinn.es/portal/site/MICINN/menuite>

⁸ Source U-Multirank : http://ec.europa.eu/education/higher-education/doc/multirank_en.pdf

Proposition de poursuite de l'adaptation de la notation à la structuration des unités de recherche (en axes, thèmes, projets, ou équipes)

La notation séparée des différentes parties d'une unité de recherche, qu'il s'agisse d'équipes, d'axes, de projets ou de thèmes, a été adoptée dans la majorité des domaines de recherche. Il apparaît en effet que l'organisation des unités de recherche révèle une hétérogénéité des composantes internes, tant en termes de positionnement thématique ou disciplinaire, et d'effectifs ou de moyens, qu'en termes de missions et productions, cette hétérogénéité étant assumée et gérée à l'échelle de l'unité et de ses tutelles. Dans ce contexte, la granularité de l'évaluation généralement adoptée est celle de l'équipe de recherche⁹, en complément de celle réalisée pour l'ensemble de l'unité.

L'évaluation, effectuée par les comités d'experts, ne prend son sens que si elle s'applique à des ensembles suffisamment cohérents. Il est cependant important de préciser que les critères d'évaluation des équipes ne sont pas automatiquement les mêmes que ceux retenus à l'échelle de l'unité, et qu'ils pourront varier selon l'équipe - en prenant en considération ses missions au sein de l'unité et ses productions. Dans la mesure où un nombre croissant d'unités s'organisent sur le mode « projet », la granularité de l'analyse menée par les experts de l'AERES doit prendre aussi en compte ces nouvelles organisations.

Evaluer les équipes, puis l'unité, amène ainsi à identifier, pour chacune de ces entités, les points forts et ceux qui doivent être confortés. Cette évaluation rend quelquefois difficile la position des équipes qui sont les moins bien évaluées au sein de l'unité. Elle doit être justifiée avec soin, mais elle s'avère utile à la direction des unités de recherche et aux directions d'établissement, pour mettre en œuvre une stratégie visant à amener l'ensemble de l'unité au meilleur niveau.

Il n'est pas proposé de modification significative de l'adaptation de la notation à la structuration des unités au cours du futur cycle d'évaluation. Selon la méthodologie actuellement en cours, la granularité d'analyse sera adaptée au fonctionnement de l'unité de recherche : équipes ou non ; organisation ou non en mode projet ou matricielle.

Par contre, la réflexion à engager sur les critères d'évaluation (§7.1) cherchera à identifier les critères facultatifs éventuellement pertinents pour le grain « équipe », en particulier sur le critère « vie de l'équipe », et en vue de l'adaptation de la grille en cas d'organisation en mode « projet ».

Propositions d'évolution concernant la note globale

A la suite des travaux préparatoires à la production de ce document et, compte tenu des points de vue exprimés au sein du conseil, il apparaît d'abord important de tenir compte du fait que les différentes catégories de destinataires des résultats de l'évaluation ont des attentes diverses vis-à-vis de l'évaluation et de la notation.

- Le premier cercle (les « évalués ») et le deuxième cercle (les décideurs/financeurs) des destinataires recevront, comme principal **produit final de l'évaluation, le rapport d'évaluation et les notes associées à chacun des critères**. Celles-ci prendront la forme d'un profil de qualité, selon la note attribuée pour chacun des critères retenus (ces derniers étant susceptibles d'évoluer à l'issue du travail sur la révision des critères) : par exemple, AABA, si 4 critères restent retenus. Cette production sera accompagnée par une fiche explicative rappelant la signification des différents critères utilisés. La fiche sera destinée, en particulier, aux décideurs/financeurs, afin qu'ils puissent utiliser les notes associées à chacun des critères à bon escient et, le cas échéant, qu'ils puissent pondérer les résultats de la notation conformément à leur stratégie.
- Les destinataires du troisième cercle, c'est-à-dire ceux qui ne prennent pas de décision directe, notamment financière, pourront avoir accès, comme **produit final de l'évaluation, aussi aux deux éléments précédents, mais surtout à une appréciation d'ensemble portée sur l'unité**.

Plutôt que présentée sous la forme actuelle de la note globale A+, A, B, ou C, cette appréciation d'ensemble visera une amélioration, sous trois angles :

- éviter une confusion avec l'échelle de notation retenue pour les différents critères. Dans ce but, il ne sera plus fait appel, pour l'appréciation d'ensemble, aux lettres A+, A, B, C, qui sont utilisées pour les notes associées à chacun des critères ;
- faire en sorte que cette appréciation d'ensemble, qui résulte du travail de concertation entre les présidents des comités de visite dans un même domaine disciplinaire, soit la plus simple possible, tout en exprimant clairement des éléments qualitatifs ;
- faire en sorte aussi que cette appréciation d'ensemble ne risque pas de se substituer, comme c'est parfois le cas actuellement avec la note globale, à la richesse apportée par des notes associées à des critères diversifiés et complémentaires, dans une vision large des produits de la recherche.

⁹ L'« équipe » désigne ici tout sous-ensemble d'une unité de recherche, correspondant à un groupe de chercheurs travaillant dans un sous-domaine donné. L'« équipe » comporte le plus souvent entre 3 et 15 personnes. Dans certains domaines d'activité, le sous-ensemble d'une unité de recherche peut porter le nom d'« axe », de « thème » ou de « projet de recherche ».

Selon ces buts, l'appréciation d'ensemble, produite collégialement par les pairs informés, afin de « vulgariser », autant que possible, le résultat de leur travail d'évaluation, prendra la forme d'une appréciation textuelle synthétique soulignant les faits marquants, à partir d'une série de commentaires préétablis reliés aux différentes combinaisons possibles de la notation multicritères ; cette appréciation sera cohérente avec le souci, durant la seconde étape de la phase d'évaluation des unités de recherche, de maintenir au premier plan une évaluation par les pairs, qualitative et contextualisée ;

Cette forme apparaît préférable à une cotation par étoilement, avec trois, quatre ou cinq niveaux. Si le passage à l'étoilement pourrait être vu comme un progrès éditorial par rapport à la note globale actuelle, il nécessiterait néanmoins : (1) de bien expliciter les limites d'usage de l'étoilement par rapport à la notation multicritères ; (2) de décrire précisément l'opération qui fait passer du système de notation multicritères à l'étoilement : est-ce une sorte de composite ? Est-ce une forme de normalisation ? Est-il valable tous domaines confondus ? Etc.).

La proposition est qu'une appréciation d'ensemble se substitue à la note globale A+, A, B, C. Cette appréciation d'ensemble sera issue du travail collégial des pairs et visera à condenser le résultat de leur travail d'évaluation, sous une forme accessible à un public large, notamment aux destinataires du troisième cercle. Elle prendra la forme d'un court texte (moins de 180 caractères), exprimant de manière littérale simple le message principal contenu dans le rapport d'évaluation et, en particulier, les notes multicritères.

L'appréciation littérale sera établie sur la base d'une série de commentaires préétablis correspondant à chacune des combinaisons de notes multicritères. Ce message pourra éventuellement être retouché, en accord avec le(s) délégué(s) scientifique(s) chargés de l'évaluation de l'unité, par le président du comité d'experts qui a procédé à la visite de l'unité, de manière à en adapter le contenu à la situation spécifique de l'unité en question.

Le maintien de cette appréciation d'ensemble, qu'elle qu'en soit la forme, permettra aussi à l'AERES d'affirmer sa volonté de prendre l'ensemble de ses responsabilités en matière d'évaluation, notamment celle de simplifier autant que possible, à l'intention de certains destinataires, le résultat de celle-ci, et de donner toute son ampleur à l'évaluation par les pairs, sans déporter une part de sa responsabilité sur les utilisateurs de l'évaluation.

Ce maintien visera aussi à limiter le risque qu'une notation globale « sauvage » soit mise en place.

Proposition de maintien concernant la distinction de l'excellence

Le maintien, en l'état, de la notation multicritères laissera la capacité de conférer une signification d'excellence à la note A+, sur tel ou tel des critères d'évaluation. Il sera toutefois important de souligner, dans la fiche explicative de la notation multicritères, que l'excellence d'une unité de recherche est difficilement dissociable de celle des personnes la composant. La notion d'excellence prend un meilleur sens à l'échelle individuelle, pouvant être identifiée par différents moyens et avec différents niveaux possibles¹⁰. A l'échelle d'un site, une appréciation de l'excellence est aussi possible par assemblage des données propres aux différents chercheurs.

Proposition de renoncement à une normalisation de la distribution des niveaux de notation

Une normalisation de la distribution des niveaux de notations aurait l'avantage, dans le cadre de la notation multicritères, de préciser la signification des différents niveaux. En cas de quota faible de notes A+, elle guiderait aussi plus fortement l'identification de l'excellence.

Toutefois, compte tenu de l'objectif premier de l'évaluation, qui est de permettre à l'ensemble des unités évaluées de s'engager dans un processus d'amélioration, et en l'absence de critères nationaux ou internationaux permettant de fixer ces quotas, il apparaît préférable de renoncer à une telle normalisation.

En revanche, un suivi de l'évolution de la distribution des niveaux de notation par domaine scientifique sera mis en place et rendu public.

Proposition pour limiter le risque du « tatouage » injustifié

Un travail sera engagé pour définir les compléments d'information susceptibles d'être fournis par les unités de recherche et les conditions dans lesquelles ces éléments pourraient être rendus publics dans les mois ou années suivant l'évaluation.

¹⁰ Prix Nobel ou équivalents, réception par l'Académie des sciences ou équivalents, prix internationaux, nomination à l'Institut universitaire de France, prix nationaux....

8 - Les démarches exploratoires déjà engagées quant à une éventuelle évolution du dispositif de notation

Compte tenu du délai de quelques mois seulement restant à courir avant la mise en œuvre opérationnelle du dispositif de notation pour les cinq ans à venir et compte tenu de la nécessité d'un avis à brève échéance du conseil de l'agence sur ce sujet, une première démarche exploratoire a été engagée. Elle a confirmé l'attente de toutes les parties concernées (responsables d'organisme de recherche ou d'établissement d'enseignement supérieur et de recherche ; responsables du ministère de l'enseignement supérieur et de la recherche) vis-à-vis d'un système de notation amélioré dans sa lisibilité et sa cohérence, et encore plus ouvert à une approche qualitative.

9 - Communication

Une des leçons à tirer de l'expérience passée est qu'une part importante de l'expression d'insatisfaction, vis-à-vis de la notation, provient de l'usage qui est fait de la notation. C'est pourquoi une initiative importante, dans le cadre de l'accompagnement du nouveau dispositif de notation, sera la mise en place d'un programme de communication. Ce programme visera à renforcer l'alignement de la gouvernance de la notation (notation multicritères, profil de qualité, appréciation d'ensemble) sur les trois grands objectifs de l'évaluation (chapitre 4) et sur les trois grandes catégories de destinataires identifiées (chapitre 5).

La communication à l'intention des unités évaluées s'appuiera sur la diffusion, à l'intention de l'ensemble des responsables d'unité de recherche, du document expliquant les nouvelles dispositions et les raisons de son évolution par rapport au dispositif actuel.

La communication à l'intention des décideurs/financeurs, c'est-à-dire des responsables de département ministériel, d'organisme de recherche, d'établissement d'enseignement supérieur et de recherche, de conseil régional, d'association de soutien à la recherche, etc., s'appuiera sur la diffusion du présent document, puis ultérieurement d'une fiche explicative de la notation multicritères.

La communication à l'intention, notamment, du grand public et des représentants élus, s'appuiera sur des relais presse, en vue d'expliquer l'évolution de l'évaluation des unités de recherche, et sur une mise à jour du site de l'AERES. Il est en effet prioritaire qu'un effort important de transparence soit produit, afin que les utilisateurs au sens large disposent d'éléments d'information aussi compréhensibles que possible sur l'évaluation et la notation des unités de recherche, et qu'ils puissent en connaître le mode d'élaboration et les limites d'utilisation.

10 - Calendrier de mise en œuvre et travaux complémentaires

La mise en œuvre d'un nouveau dispositif de notation pour le cycle quinquennal à venir sera précédée d'une phase d'échange avec les parties prenantes.

La principale proposition concernant le nouveau dispositif (voir au point 7.3) sera opérationnelle au 15 février 2012.

Des travaux complémentaires seront engagés sans délai :

- la mise en place d'un groupe de travail pour réviser, en vue d'amélioration, les critères de la notation ;
- l'élaboration d'une fiche explicative décrivant les méthodes et les différents critères utilisés pour la notation multicritères (incluant éventuellement de nouveaux critères, selon les conclusions du groupe de travail évoqué ci-dessus) ;
- l'élaboration des éléments de communications ;
- l'élaboration d'un outil informatique permettant une notation multicritères personnalisée en fonction des choix de l'utilisateur ;
- la mise en place d'un suivi de l'évolution de la distribution des niveaux de notation par domaine scientifique, ainsi que des impacts à distance de l'évaluation sur les unités et équipes évaluées ;
- les travaux visant à minimiser le risque que l'évaluation d'une unité de recherche devienne un « tatouage » injustifié.

La mise en œuvre du nouveau dispositif n'interdira pas la mise en route parallèle d'une expérimentation destinée à tester d'éventuelles améliorations de la notation pour l'avenir, lorsque de telles améliorations potentielles sont identifiées.

ANNEXE 1 : Modalités d'évaluation des activités de recherche dans différents pays européens

Organisme (Pays) Procédure	Entités évaluées	Dimensions examinées	Critères ou indicateurs	Échelle de notation	Spécificités
CHE (Allemagne) http://www.che-concept.de	Universités, sur la base d'une analyse par domaine disciplinaire.	9 dimensions : <ul style="list-style-type: none"> • Environnement (qualité infrastructure : ville, université), • Employabilité, carrières, • Equipements, ressources, • Caractéristiques des étudiants, • Performance des étudiants, • Internationalisation, • Formations académiques, • Recherche, • Opinions globales des étudiants et des enseignants, 	34 indicateurs sur ces 9 dimensions	Indique, pour chaque indicateur, le groupe dans lequel se situe l'établissement (département) : « <i>top, middle, bottom</i> »)	<p>Un classement multi-dimensions par domaine disciplinaire pour accompagner le choix des étudiants.</p> <p>Un classement personnalisable par l'utilisateur en fonction du choix des modules et des domaines (« <i>subjects</i> »).</p> <p>CHE réalise également un classement « compact » des universités uniquement sur une base disciplinaire en trois groupes : « <i>top, middle, bottom</i> » et sur un nombre restreint de critères. Ces critères sont spécifiques d'un domaine.</p>
Projet UE U-Multirank (CHERPA consortium, Allemagne, Pays-Bas, France) http://ec.europa.eu/education/highereducation/doc/multirank_en.pdf	Universités ou domaines scientifiques	5 dimensions : <ul style="list-style-type: none"> • Enseignement, et formations, • Recherche, • Transfert de connaissance, • Internationalisation, • Engagement régional. 	Au premier niveau, 15 indicateurs regroupés en 5 dimensions	vert, orange, rouge	<p>Un classement multi-dimensions d'entités qui présentent des caractéristiques (profil d'activités) comparables.</p> <p>Un profil de performance par dimension adossé à un jeu d'indicateurs pour chaque dimension.</p> <p>Un classement personnalisable par l'utilisateur en fonction des dimensions et indicateurs de son choix.</p>

Organisme (Pays) Procédure	Entités évaluées	Dimensions examinées	Critères ou indicateurs	Échelle de notation	Spécificités
<p>HEFCE (UK)</p> <p>REF (Research excellence framework)</p> <p>http://www.hefce.ac.uk/research/ref/</p>	<p>Universités</p> <p>Le grain d'analyse est celui d'un groupe de chercheurs (<i>unit</i>) d'un domaine disciplinaire tel que défini par les périmètres des panels d'évaluation (« <i>unit of assessment</i> »)</p>	<p>3 dimensions :</p> <ul style="list-style-type: none"> • les produits de la recherche, • l'impact de la recherche • l'environnement de recherche 	<p>Évaluation qualitative par les pairs sur chacun des produits et éléments associés à chacune des trois dimensions :</p> <ul style="list-style-type: none"> • Les productions des scientifiques (limitées à 4 productions majeures par scientifique) • l'impact socio-économique de la recherche portée par l'unité (une déclaration générale et des études de cas) • l'environnement (les infrastructures, l'organisation, la formation par la recherche, les ressources, la stratégie, ...) 	<p>Un profil qualité de chacune des dimensions évaluées qui indique la proportion de l'activité évaluée que l'on classe dans chacun des 5 niveaux de * à **** et non classé (« <i>sub-profile</i> »)</p> <p>et</p> <p>Un profil qualité global qui intègre les 3 dimensions après pondération (Recherche 65% ; impact 20% ; environnement 15%) (« <i>overall quality profile</i> »)</p>	<p>Le nombre et la composition des groupes dont les productions sont soumises à l'évaluation est de la responsabilité de l'établissement.</p> <p>Une entité peut demander une évaluation trans-panel.</p> <p>Une évaluation portant uniquement sur le bilan.</p> <p>Des critères d'évaluation adaptés par panel. (NB : début janvier 2012 des compléments sur les données précises requise pour la soumission et sur les critères par panel.)</p> <p>Diffusion publique du profil global.</p>

Organisme (Pays) Procédure	Entités évaluées	Dimensions examinées	Critères ou indicateurs	Échelle de notation	Spécificités
<p>KNAW-VSNU-NWO (Pays-Bas)</p> <p>SEP (Standard evaluation protocol)</p> <p>http://www.know.nl/content/Internet_KNAW/publicaties/pdf/20091052.pdf</p>	Instituts et groupes de recherche (ou programmes)	<p>3 dimensions :</p> <ul style="list-style-type: none"> • Résultats académiques • Résultats pour la société • Formation des prochaines générations de chercheurs 	<p>Appréciation qualitative par les pairs sur la base de 10 ou 11 sous-critères regroupés en 4 critères principaux :</p> <ul style="list-style-type: none"> • qualité • productivité • pertinence sociétale • vitalité et faisabilité 	<p>Une note de 1 à 5 sur chacun des sous-critères</p> <p>La notation est réalisée par le comité de visite.</p>	<p>Une attention affirmée aux spécificités des champs disciplinaires et aux missions de l'entité évaluée.</p> <p>Une évaluation rétrospective et prospective qui vise à renforcer la réflexivité des entités et à proposer des pistes d'amélioration pour l'entité, son leader ou sa hiérarchie.</p> <p>Des termes de référence produits par l'instance de gouvernance de l'institution («<i>board</i>»).</p> <p>La possibilité d'une évaluation à mi-contrat (3 ans) de la prise en compte des recommandations formulées.</p> <p>Seul « protocole » à</p> <ul style="list-style-type: none"> • Inclure une visite sur place et éventuellement des interviews de «<i>stake-holders</i>», • Faire référence explicitement aux missions spécifiques de l'entité évaluée, • Noter l'analyse SWOT effectuée par l'entité évaluée. <p>Diffusion publique des résultats de l'évaluation après une présentation du rapport au «<i>board</i>» de l'institution. Celui-ci a un an pour rendre public son avis sur les sorties et recommandations, voire leur mise en œuvre.</p>

Organisme (Pays) Procédure	Entités évaluées	Dimensions examinées	Critères ou indicateurs	Échelle de notation	Spécificités
<p>ANEP (Agencia Nacional Evaluación y Prospectiva) (Espagne)</p> <p>http://www.micinn.es/portal/site/MICINN/menuite</p>	<ul style="list-style-type: none"> • Unités, • Equipes, • Projets de recherche pour les programmes et projets du plan national • Suivi des résultats 	<p>3-4 dimensions en fonction du domaine</p> <ul style="list-style-type: none"> • Evaluation du chercheur principal • Evaluation de l'équipe de recherche • Evaluation de la qualité scientifique et de la viabilité du projet 	<p>Evaluations par les pairs (peer review) Au moins 2 évaluations par entité évaluée</p> <p>A partir des évaluations, l'équipe de coordination élabore un rapport final.</p> <p>Quand les évaluations doivent être comparatives, des commissions d'experts établissent des critères avec un ordre de priorité</p>	<p>% en fonction de chaque domaine (humanités, sciences sociales, santé, sciences expérimentales et technologiques). Ces domaines sont répartis en 26 sections.</p> <p>Ex en humanités : Projet 60% Chercheur et équipe 30% Gouvernance et financement 10%</p>	<p>La classification prétend refléter la situation actuelle des activités des centres de recherche et des universités espagnoles.</p>