

| | |
|------------|-----------------------------------------------------------------------------------------|
| Livrable 2 | Proposition d'un modèle partageable de mise en œuvre de l'évaluation qualitative à l'UB |
|------------|-----------------------------------------------------------------------------------------|

Synthèse des travaux du groupe de travail enseignants sur la mise en œuvre de l'évaluation qualitative

Participants. Anna Barry (In'Pact), Achille Braquelaire (In'Pact), Dany Carlier-Larregaray (ST-chimie), Laurence Chérel (IUT- Tech de Co), Anne-Laure Damongeot (SH-dlc), Frédéric Gaschet (DSPEG-AES), Manuel Gaudon (ST-chimie), Nathalie Gaussier (DSPEG-éco gest.), Adrien Gintrand (IUT-GCCD), Michel Hernould (ST-bio agro), Marthe-Aline Jutand (SH- sc.ed.), Carole Lemée (SH-anthr.), Catherine Madrid (IUT- Tech de Co), Laurence Maillé (ST-chimie), André Meunié (DSPEG-éco gest.), Joëlle Perroton (SH-socio), Emmanuelle Redon (SH-fc), Christophe Roiné (In'Pact), Valérie Schurdi-Levraud (ST-bio Agro), Olivier Toulemonde (ST-chimie).

Résumé. Ce document synthétise différentes propositions et remarques issues des réflexions du groupe de travail enseignants sur la mise en œuvre de l'évaluation qualitative. Ces réflexions se sont appuyées sur l'étude de la littérature et des principaux systèmes d'évaluation à l'international (voir livrable 1) et l'expérience personnelle des participants. Elles ont tout d'abord permis de conforter les objectifs visés par une amélioration du système d'évaluation des acquis, de réaffirmer le lien fort entre l'évaluation qualitative et l'alignement pédagogique et de s'accorder sur un vocabulaire et des principes communs. Elles ont également permis de préciser le rôle central des critères et grilles critériées dans l'ensemble du processus d'évaluation.

La proposition de modèle d'évaluation qualitative retenue repose sur cinq niveaux exprimant trois niveaux différents de réussite, un niveau insuffisant mais améliorable et enfin l'échec. Ces niveaux sont associés à des notes exprimées sous forme de lettres, les trois niveaux de réussite étant sous certaines conditions modulables.

Ce document décrit également pourquoi et comment utiliser un système de moyenne de points de note pour agréger un ensemble de notes afin de calculer la note finale à une UE ou la moyenne à un programme de formation, et ceci de manière suffisamment précise et fiable pour permettre, lorsque c'est demandé, un classement des diplômés. Ce système est comparé avec le système de moyenne sur 20 du point de vue de la pertinence et de la précision du résultat obtenu.

Enfin, la dernière section est consacrée à la mise en œuvre progressive de ce modèle qualitatif au sein du système quantitatif de notation sur 20 et aux possibilités et limites de la coexistence de ces deux systèmes.

1 Principes et terminologie

1.1 Objectifs retenus

Les échanges entre les participants du groupe de travail enseignants du projet EQAE ont confirmé l'intérêt partagé pour expérimenter un modèle d'évaluation qui permette d'**améliorer le retour fait aux étudiants sur la qualité de leurs apprentissages**. Ce retour doit permettre à un étudiant de comprendre ce qui ne va pas dans le travail qu'il a remis, de savoir ce qui a manqué, et de pouvoir ensuite **construire son apprentissage de manière plus individualisée et pertinente**.

Le premier objectif visé est donc de pouvoir, indépendamment de l'attribution d'une note, **mettre des mots sur ce que les étudiants sont capables de faire** et définir à partir de quel moment on considère qu'ils ont atteint les objectifs de l'enseignement. Autrement dit, il ne s'agit pas d'attribuer simplement une note, mais de pouvoir exprimer, en fonction de ce que l'étudiant a produit, ce qu'il sait faire, dans quel contexte, avec quelles ressources et dans quelles conditions d'autonomie.

Un deuxième objectif est de pouvoir **valider la réussite à un programme d'études** également de manière qualitative, et d'**améliorer la lisibilité du diplôme** en termes de poursuite d'études ou d'employabilité.

Ce sont ces deux premiers objectifs que l'on résume dans la suite de ce document au moyen du terme générique d'« **évaluation qualitative** ».

Un troisième objectif est, le cas échéant, de permettre d'**effectuer un classement** entre les diplômés d'une même promotion.

En résumé, il s'agit de proposer un modèle d'évaluation des acquisitions des étudiants permettant de satisfaire conjointement aux objectifs d'une évaluation formative et d'une évaluation sommative, les bénéfices attendus étant :

- une **amélioration des apprentissages** en permettant aux étudiants de mieux appréhender leurs acquisitions et leurs marges de progression ;
- une **amélioration de la cohérence des notes produites** pour qualifier les résultats obtenus par les étudiants aux activités d'évaluation ;
- une amélioration de la **personnalisation de la progression des étudiants**, davantage en adéquation avec la réalité de leurs acquisitions ;
- un renforcement et une **meilleure maîtrise de l'approche par compétences**, grâce à une appréciation plus qualitative des apprentissages.

D'un point de vue pratique, l'évaluation qualitative d'un enseignement doit pouvoir être mise en œuvre de manière progressive au sein d'un programme de formation, ce qui suppose de pouvoir faire coexister des évaluations qualitatives et quantitatives (voir livrable 1), *a minima* durant une période transitoire.

1.2 Évaluation qualitative et alignement pédagogique

La mise en œuvre de l'évaluation qualitative est indissociable de celle du principe d'**alignement pédagogique** pour lequel les **acquis d'apprentissage (AA)**, ou *learning outcomes*, jouent un **rôle central**. On peut résumer ce principe par les quatre étapes suivantes :

1. décrire les objectifs d'un enseignement sous la forme de résultats attendus en termes d'apprentissage : **acquis d'apprentissage visés (AAV)** et niveau que les étudiants doivent atteindre ;
2. créer un **environnement d'apprentissage** à l'aide d'activités d'enseignement/apprentissage en cohérence avec les AAV ;
3. utiliser des **activités d'évaluation (AE)** également en cohérence avec les AAV et apprécier de quelle manière les performances des étudiants répondent aux résultats attendus ;
4. transformer ces jugements en **critères standards de notation**.

Chaque AAV est exprimé au moyen d'un verbe décrivant l'activité d'apprentissage et d'un objet indiquant son contenu. Il s'accompagne de critères explicitant les différents niveaux d'apprentissage. La cohérence entre les différentes étapes résulte de l'utilisation des mêmes termes pour décrire les acquis d'apprentissage visés, les activités d'apprentissage et les activités d'évaluation. En particulier les critères d'appréciation des performances des étudiants doivent traduire de quelle manière l'AAV est maîtrisé dans le cadre de l'AE.

L'évaluation qualitative d'un étudiant concerne directement les étapes 1, 3 et 4. Ces différentes étapes sont souvent prises en compte par les enseignants de manière implicite, mais elles sont plus rarement explicitées et partagées avec les étudiants. La mise en œuvre de l'évaluation qualitative doit donc être **l'occasion d'une mise en œuvre plus rigoureuse et maîtrisée de l'alignement pédagogique** par les enseignants, notamment en portant une attention particulière à l'explicitation des différentes étapes et à leur partage avec les étudiants.

1.3 Terminologie et principes généraux

À cette étape de la réflexion, on propose d'adopter la terminologie et les principes généraux qui suivent.

Critères, indicateurs et niveaux d'apprentissage. Les **critères** explicitent les objectifs pédagogiques au moyen d'indices observables. Pour chaque critère, la qualité des apprentissages est décrite par des **indicateurs** qui explicitent une hiérarchie de **niveaux d'apprentissage** et la manière d'apprécier si un niveau a été ou non atteint. On peut formuler les observations suivantes :

- une hiérarchie de niveaux peut être décrite de manière **générique**, pour qualifier la manière de réussir à une UE ou un programme de formation dans son ensemble, ou de manière plus **spécifique**, relativement à une AE ;
- les niveaux génériques et spécifiques doivent être exprimés de manière **cohérente** les uns par rapport aux autres : le niveau spécifique est une formulation opérationnelle du niveau générique à l'échelle d'une AE ;
- le niveau générique doit être **largement partagé** afin de permettre d'agrèger des notes de différentes activités ou UE d'une manière qui préserve la dimension qualitative des évaluations.

Activité d'évaluation. Une **activité d'évaluation** (AE) désigne tout type d'activité utilisée pour apprécier un niveau d'apprentissage par référence à des critères et des indicateurs. On peut formuler les observations suivantes :

- une AE peut indifféremment être un exposé, la production d'un dossier ou d'un mémoire, un devoir (examen sur table, devoir à domicile), un QCM, etc. (à ce niveau, la forme et la modalité retenues n'entrent pas en ligne de compte) ;
- pour une AE donnée, il convient de formuler l'ensemble des critères décrivant les objectifs de l'activité et des indicateurs explicitant la hiérarchie des niveaux d'acquisition et la manière de les apprécier ;
- une AE peut être associée à un ou plusieurs AAV.¹

Appréciation. L'**appréciation** du niveau atteint à une AE est effectuée par référence à la hiérarchie d'indicateurs explicitant ce sur quoi porte l'activité et la caractérisation des différents niveaux de maîtrise de cette activité. On peut formuler les observations suivantes :

- l'étape d'appréciation du niveau atteint **précède** celle de l'attribution d'une note (l'appréciation d'une activité d'évaluation ne se résume pas à la production d'une note) ;
- l'appréciation est conduite de manière purement qualitative lorsqu'elle se réfère directement à une **grille critériée** pour déterminer le niveau atteint, sans recourir à des pourcentages ou des scores (par exemple comptabilisation d'un nombre de réponses correctes) ;
- l'appréciation peut également être conduite de manière quantitative, par exemple en comptabilisant les bonnes réponses à un QCM, mais le score obtenu doit alors **pouvoir être mis en correspondance avec un niveau de maîtrise exprimé qualitativement**.

1. Voir livrable 1, section 3.4.

Note. La **note** est l'expression par une lettre ou un nombre du niveau atteint par l'étudiant. Idéalement, le nombre de notes du système de notation est identique au nombre de niveaux d'apprentissage critérisés et appréciables.

Évaluation. Dans le contexte de l'évaluation qualitative, le terme d'évaluation désigne le **processus complet mis en œuvre pour apprécier et qualifier le niveau d'apprentissage d'un étudiant**, c'est-à-dire :

- la définition des critères et des indicateurs d'appréciation du niveau atteint,
- la définition des activités d'évaluation et l'organisation de ces activités,
- l'appréciation du niveau atteint,
- l'attribution d'une note traduisant le résultat de cette appréciation.

Agrégation de notes. L'**agrégation** de notes consiste à combiner plusieurs notes exprimant un résultat partiel afin de produire une note plus globale. À un niveau micro, cela permet de combiner des notes correspondant à plusieurs sous-activités au sein d'une AE pour produire l'appréciation et la note du résultat à l'AE. À un niveau plus large, cela permet de combiner les notes de plusieurs activités pour évaluer une UE (voire un groupe d'UE). À un niveau plus large encore, cela permet de combiner les notes finales de tous les enseignements d'un programme de formation pour déterminer la réussite à ce programme. On peut formuler les observations suivantes :

- l'agrégation des notes doit traduire une agrégation des niveaux d'appréciation en respectant les hiérarchies de niveaux, ce qui implique de rechercher une cohérence entre l'expression micro et macro des niveaux d'apprentissage ;
- la prise en compte des différentes notes à agréger doit pouvoir être pondérée en fonction de l'importance relative des sous-activités, activités ou enseignements évalués.

Remarque 1. L'agrégation d'un ensemble de notes peut exprimer un résultat plus fin que le nombre de niveaux, et donc que le nombre de notes utilisées. Cela peut être traduit par une modulation autour de la note indiquant la manière dont le niveau a été atteint, par exemple entre « à peine satisfait » et « satisfait de manière remarquable », mais cela ne correspond en aucun cas à l'introduction de nouveaux niveaux intermédiaires, ni par conséquent à celle de nouveaux critères. Cette modulation peut éventuellement être également utilisée au moment de l'évaluation lorsque l'AE mise en œuvre s'y prête.

2 Proposition d'un modèle

2.1 Choix d'un nombre de niveaux

Le benchmark des systèmes d'évaluation dans l'enseignement supérieur présenté dans le livrable 1 montre qu'il existe, d'un pays à l'autre, une relative diversité du nombre de niveaux d'appréciation des résultats, pouvant aller de trois à dix², mais avec une forte prédominance des systèmes à quatre ou cinq niveaux. On rencontre également parfois des variations entre les différentes universités d'un même pays.³ En revanche, ce nombre de niveaux est généralement unique au sein d'une même université. Cette condition semble notamment incontournable pour développer les porosités entre les différents programmes de formation et permettre aux étudiants de construire des cheminements pluridisciplinaires.

Il apparaît donc nécessaire de fixer un **nombre de niveaux qui soit le même pour l'ensemble de l'établissement** et de partager une **hiérarchie commune de critères génériques** pour décrire ces niveaux à l'échelle macro des programmes de formation, si l'on veut faciliter et encourager la **mutualisation** d'enseignements⁴ et mettre en œuvre un **processus pertinent** d'agrégation de notes.

2. Si l'on fait abstraction des quelques systèmes quantitatifs reposant sur une notation sur 20 (France, Belgique, Portugal et pays du Maghreb pour l'essentiel).

3. Australie ou Canada par exemple.

4. UE d'ouverture par exemple

On constate enfin à l'échelle internationale que pour les systèmes majoritaires d'évaluation reposant sur quatre ou cinq niveaux, le dernier niveau se distingue généralement des autres par le fait qu'il n'est pas associé à un critère explicite, mais qu'il traduit simplement l'échec à atteindre le critère minimal.⁵ Cela correspond par conséquent à des systèmes respectivement à trois niveaux plus un et à quatre niveaux plus un.

Nous proposons de retenir un système à **cinq niveaux**, c'est-à-dire quatre niveaux critérisés plus un, organisé selon le principe suivant :

- les quatre niveaux critérisés sont décrits par des indicateurs génériques et partageables à l'échelle de l'établissement ;
- les trois premiers niveaux traduisent différents **niveaux de réussite** :
 - **suffisant** : un minimum d'éléments attendus pour permettre une validation des apprentissages est présent et ces éléments sont formulés de manière correcte ;
 - **bon** : tous les éléments attendus sont présents et formulés de manière correcte ;
 - **excellent** : la réponse est précise, complète et traduit une maîtrise de l'ensemble des acquisitions visées.
- le quatrième niveau correspond à une réponse partielle qui doit être améliorée pour atteindre le niveau suffisant pour une validation :
 - **insuffisant** : la réponse contient des éléments attendus mais n'atteint pas le niveau minimum requis.
- le cinquième niveau (**échec**) est défini implicitement, sans indicateur explicite ; il traduit une réponse qui ne permet même pas d'atteindre le niveau insuffisant : absence d'éléments pertinents dans le travail remis, plagiat, etc.

Remarque 2. Le groupe de travail a débattu de l'intérêt de conserver ou non deux niveaux (insuffisant et échec) pour traduire une absence de réussite. La proposition de les conserver a été retenue pour deux raisons :

- cela permet d'encourager un étudiant à travailler pour atteindre le niveau suffisant dans le cas d'une évaluation continue ;
- cela permet le cas échéant de différencier un échec compensable sous certaines conditions d'un qui ne le serait pas.

Remarque 3. Le groupe de travail a également débattu de la pertinence de pouvoir ou non exprimer des modulations autour d'un niveau atteint. Cette modulation peut résulter de l'agrégation d'un ensemble de notes ou traduire le cas échéant une appréciation subjective par l'évaluateur de la manière dont il considère que le niveau a été atteint (voir remarque 1 page 4). La proposition de conserver cette possibilité a été retenue en raison de sa pertinence pour exprimer plus finement le résultat d'une agrégation d'un ensemble de notes. Son utilisation pour l'évaluation d'une AE a fait l'objet de réserves. Si cette possibilité est utilisée, elle ne doit pas interférer avec l'appréciation du niveau atteint. En effet, cette appréciation ne devrait dépendre que des critères associés aux niveaux d'apprentissage. Dans tous les cas, il est proposé de réserver cette possibilité de modulation aux seuls niveaux de réussite.

● En résumé, le modèle proposé permet d'exprimer trois niveaux de réussite, un niveau insuffisant et un niveau d'échec. Au niveau micro, ces quatre niveaux sont explicités pour chaque AAV et/ou pour chaque AE au moyen d'une hiérarchie d'indicateurs spécifiques. Le niveau *suffisant* doit permettre d'expliciter le minimum attendu pour valider une AE, et par agrégation un enseignement. Le niveau *excellent* doit traduire une pleine maîtrise des apprentissages visés. Les autres niveaux peuvent être exprimés par référence à ces deux niveaux. La hiérarchie d'indicateurs associés à une évaluation doit permettre :

- à l'enseignant d'apprécier objectivement le niveau atteint, niveau qui se traduira par une note ;
- à l'étudiant de comprendre les points forts et les points faibles de sa réponse et d'identifier ses marges de progrès.

5. Voir livrable 1, section 2.3.

2.2 Proposition d'un système de notation

Trois critères ont été privilégiés pour traduire ce modèle par un système de notation :

- reposer sur un nombre de notes égal au nombre de niveaux ;
- permettre de différencier clairement les quatre niveaux critérisés et l'échec complet ;
- permettre de différencier clairement le niveau d'apprentissage atteint et une éventuelle modulation autour de ce niveau (voir remarque 3).

Cela conduit assez naturellement à retenir un système reposant sur des lettres

- A, B et C pour exprimer de manière décroissante les trois niveaux de réussite ;
- D pour exprimer le niveau insuffisant ;
- F pour exprimer l'échec, le choix de F et non de E permettant de marquer clairement la discontinuité entre les quatre niveaux critérisés et l'échec complet.

Ce choix conforte l'idée que l'appréciation du résultat est faite par référence à des niveaux de réussite critérisés et non à une mesure⁶, ce qui ne serait pas aussi explicitement le cas avec des notes numériques.

L'éventuelle modulation des trois niveaux de réussite est exprimée au moyen des deux symboles spécifiques + et - appliqués aux notes A, B et C : A+ et A-, B+ et B-, C+ et C-. Du point de vue méthodologique, si la modulation est utilisée pour évaluer une AE, l'étape d'appréciation du niveau de réussite à une AE (et donc l'attribution d'une lettre) précède l'étape de la modulation éventuelle de cette appréciation. Concrètement, si la hiérarchie d'indicateurs de notation d'une AE est clairement exprimée, on peut éventuellement moduler un A en A- ou A+, mais on ne devrait pas être amené à faire glisser une appréciation de A- vers B+ (soit les indicateurs permettant l'attribution d'un A sont globalement satisfaits, soit ils ne le sont pas).

Enfin ce choix correspond au système de notation majoritairement adopté à l'international, ce qui en fait *a priori* un système éprouvé.

2.3 Définition des critères et des indicateurs

L'attribution d'une note finale à un programme de formation résulte de l'agrégation des notes finales obtenues aux différentes UE de ce programme. De même l'attribution d'une note finale à une UE peut résulter de l'agrégation des notes obtenues aux différentes AE de cette UE. Enfin, une AE peut elle-même comporter plusieurs sous-activités notées séparément et agrégées pour former la note de l'activité. Cela suppose d'être capable d'effectuer de manière cohérente ces agrégations depuis le niveau micro de l'AE jusqu'au niveau macro du programme de formation.

Dans le cas où plusieurs programmes partagent différentes UE⁷, il est important que les critères d'appréciation du niveau atteint soient définis de manière cohérente entre ces différents programmes. C'est une condition nécessaire pour préserver la dimension qualitative de l'évaluation et la compatibilité des notes produites. Cette exigence de cohérence a conduit à préconiser de fixer un nombre de niveaux unique pour l'ensemble de l'établissement (cf § 2.1 page 4). Elle nécessite également qu'à l'échelle macro des programmes de formation, les différents niveaux partagent une **formulation générique commune** des indicateurs. À l'échelle intermédiaire d'une UE ou micro d'une AE, il s'agit alors de traduire cette formulation commune et partagée en indicateurs spécifiques explicitant les différents niveaux d'acquisition dans le contexte de l'UE ou de l'AE.

Remarque 4. À terme, la définition des indicateurs génériques communs du niveau macro devrait être arrêtée à l'échelle de l'établissement afin de garantir que les équipes pédagogiques qui désirent s'approprier et mettre en œuvre une évaluation qualitative puissent converger vers un système d'évaluation partagé. À l'échelle micro, la construction des critères et indicateurs est en revanche du ressort de chaque équipe pédagogique ou enseignant en charge des AE.

6. Voir livrable 1.

7. Par exemple des UE d'ouverture ou des UE mutualisées au sein de programmes pluridisciplinaires.

Remarque 5. La formulation générique commune des niveaux de réussite doit probablement prendre en compte le niveau de maturité des étudiants, que ce soit en fonction de leur cycle d'étude ou de leur degré d'avancement au sein d'un programme de formation.

Tous ces éléments ont vocation à être précisés en s'appuyant sur les retours d'expérience des premières expérimentations qui pourront être conduites.

2.4 Agrégation des notes

Pour un étudiant, l'agrégation des notes finales obtenues à chaque UE de son parcours peut se faire simplement en reprenant le système nord-américain *Grade Point Average* (GPA) qui permet de calculer une note finale à un programme de formation à partir d'une moyenne des notes finales de chaque enseignement. On attribue pour cela à chaque note un nombre de points de la manière suivante :

| Correspondance notes → points de note | | | | |
|---------------------------------------|--------------|--------------|-------------|-------------|
| A : 4 points | B : 3 points | C : 2 points | D : 1 point | F : 0 point |

Il est également possible de prendre en compte les modulations en ajoutant ou retranchant une petite valeur au nombre de points de la note selon que la modulation est positive ou négative. Nous proposons de reprendre la valeur de 0.3 utilisée dans le système GPA.⁸ Ce choix permet d'exprimer que l'écart inter-classes est supérieur à l'écart intra-classe.⁹ On calcule ensuite la **moyenne des points de note** (MPN), en pondérant au besoin ces points de note par des coefficients exprimant l'importance relative de chaque enseignement, par exemple selon son nombre de crédits.

Cette approche est **pertinente** car les écarts numériques entre les points correspondent à des écarts qualitatifs d'appréciation, ce qui n'est pas le cas avec une échelle de notes sur 20.¹⁰ La MPN permet d'exprimer deux choses différentes :

- la **note finale** à un programme de formation, qui est celle dont la valeur en points de note est la plus proche de la MPN ;¹¹
- une moyenne numérique qui, lorsque c'est souhaité, peut être utilisée pour effectuer un **classement** entre les étudiants validant simultanément un programme de formation.

Remarque 6. Dans le livrable 1 nous décrivions une méthode, adaptée de Biggs et Tang¹², pour agréger les notes de plusieurs sous-activités d'une AE et permettre une évaluation multicritère. Une comparaison de cette méthode avec la méthode MPN montre qu'elles sont équivalentes et que l'on peut donc utiliser la MPN pour effectuer toutes les agrégations, ce qui simplifie la mise en œuvre de ce processus.

2.5 Précision de la méthode

Un enseignant habitué au système de notation sur 20 et attaché au principe de classement des étudiants pourrait formuler l'objection que cette méthode manque de précision pour exprimer le résultat final.

Pourtant, les défauts bien identifiés du système de notation sur 20¹³ font que la précision apparente de la note obtenue, parfois exprimée jusqu'à deux décimales, est parfaitement illusoire. On peut même considérer que cette précision en termes de « justesse » du résultat diminue lorsque le nombre de notes agrégées augmente.

8. Par exemple C+ vaut 2.3 points et B- vaut 2.7 points.

9. Par exemple, l'écart entre C+ et B- est supérieur à l'écart entre C et C+ ou à celui entre B- et B.

10. Voir livrable 1.

11. Par exemple une MPN de 2,81 est plus proche de 2,7 que de 3 et correspond donc à la note B-, ou à la note B si on ne souhaite pas faire apparaître de modulation sur la note finale du diplôme.

12. Voir références dans le livrable 1.

13. Par exemple l'incohérence docimologique du mode de production de résultat, l'accumulation des erreurs avec l'augmentation du nombre de notes intervenant dans le calcul de la moyenne, ou encore la tendance de ce système à lisser les résultats — voir note de présentation du projet et livrable 1.

A l'inverse, la MPN conserve le caractère qualitatif de la note et de l'appréciation du travail de l'étudiant. Considérons par exemple le cas d'un étudiant ayant obtenu les quatre notes B, B, A et B. Sa MPN est de 3.25 et sa note finale est donc B+. Considérons ensuite un second étudiant ayant obtenu les notes B-, B+, A+ et B. Sa MPN est de 3.33 ce qui correspond également à B+. Si maintenant l'on souhaite classer les deux étudiants, la MPN du second étudiant le place devant le premier, et l'on peut se persuader facilement que ce résultat traduit plus fidèlement la qualité des résultats obtenus par les étudiants que ne le ferait une moyenne de notes sur 20 dont les écarts ne sont pas cohérents entre eux.

Quant à la **précision** de la MPN, elle peut être appréciée à partir du nombre de moyennes différentes que l'on obtient en combinant les différentes MPN possibles. Par exemple avec une dizaine de notes non modulées, c'est-à-dire pouvant chacune prendre les valeurs entières de 0 à 4, on obtient une quarantaine de MPN différentes, soit autant que de notes sur 20 exprimées au demi-point près. Si maintenant l'on s'autorise des modulations, et toujours avec une dizaine de notes on peut obtenir 378 MPN différentes.

Sachant que les MPN sont distribuées sur une échelle de 0 à 4.3 cohérente du point de vue de l'appréciation qualitative des résultats, on peut considérer que **la précision est non seulement suffisante pour effectuer un classement au sein d'une promotion, mais qu'elle est même supérieure à celle de la moyenne d'un même nombre de notes sur 20 attribuées sans critère qualitatif.**

3 Possibilités de mise en œuvre progressive

La mise en œuvre progressive de l'évaluation qualitative au sein d'un programme de formation peut être considérée de deux façons différentes. Si la pratique majoritaire au sein du programme de formation demeure quantitative, il s'agit d'intégrer les notes qualitative à l'ensemble des autres notes, et donc de **convertir des notes d'UE produites qualitativement en notes sur 20 ayant approximativement la même signification.**

Pour cela il est proposé d'associer les trois niveaux de réussite à l'intervalle de notes sur 20 allant de 10/20 à 20/20, le niveau « *insuffisant* » à l'intervalle allant de 8/20 à 10/20 (exclu), et enfin le niveau « *échec* » à l'intervalle allant de 0/20 à 8/20 (exclu). Cette correspondance est résumée sur le tableau suivant :

| niveaux de réussite | A, B et C | [10/20, 20/20] |
|---------------------|-----------|-----------------|
| <i>insuffisant</i> | D | [8/20, 10/20 [|
| <i>échec</i> | F | [0/20, 8/20 [|

Concrètement cela revient à mettre les MPN en relation avec les notes sur 20 de la manière suivante :

| | | |
|-------------|---|-----------|
| [0, 0.5] | → | [0, 8] |
| [0.5, 1.7] | → | [8, 10] |
| [1.7, 4.3] | → | [10, 20] |

Si au contraire la pratique majoritaire au sein d'un programme de formation est devenue qualitative, il s'agit de manière symétrique de convertir des notes produites quantitativement et exprimées sur 20 en l'un des cinq niveaux de réussite, le cas échéant modulé. Comme la notation sur 20 n'est pas critérisée, il ne peut y avoir de solution pleinement satisfaisante du point de vue pédagogique. La solution la plus acceptable est *a priori* de mettre en œuvre l'inverse de la transformation précédente pour calculer une valeur de points de note à partir de la note sur vingt.¹⁴

14. On aurait pu également construire une correspondance reposant sur le système français de mention (TB, B, AB et P) mais cela reviendrait à faire artificiellement correspondre un système quatre niveaux avec les trois niveaux d'appréciation de réussite A, B et C, et de toute façon, le système français de mentions n'est pas un système qualitatif construit sur des critères objectivés. Ce principe est utilisé dans les pays du Maghreb en faisant correspondre les quatre mentions du système français aux notes C, B, A et A+, ce qui revient à considérer que les deux niveaux B et TB ne sont pas véritablement différenciés et ne sont que de simples modulations au sein d'un même niveau.

Remarque 7. Les notes qualitatives sont associées à des intervalles de notes sur 20 dont les amplitudes varient dans un rapport de un à quatre.¹⁵ Cette irrégularité ne fait que traduire l'inadéquation de l'échelle de la notation sur 20 pour mettre en œuvre une évaluation qualitative.

Remarque 8. La correspondance proposée présente des faiblesses du point de vue de l'évaluation qualitative. Elle a cependant le mérite de permettre de faire coexister les deux systèmes qualitatif et quantitatif au sein d'un même programme de formation. On peut juger de l'acceptabilité de cette proposition en la construisant pour les différentes notes avec les modulations éventuelles :

| note : | A+ | A | A- | B+ | B | B- | C+ | C | C- | D | F |
|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|
| points de note : | 4.30 | 4.00 | 3.70 | 3.30 | 3.00 | 2.70 | 2.30 | 2.00 | 1.70 | 1.00 | 0.00 |
| correspondance /20 : | 20.00 | 18.85 | 17.69 | 16.15 | 15.00 | 13.85 | 12.31 | 11.15 | 10.00 | 8.83 | 0.00 |

Bien entendu le calcul de la moyenne des points de note peut produire des valeurs intermédiaires aux points de note. Par exemple une MPN de 3.5 correspond à A- et 16.9/20 et une MPN de 0.3 correspond à F et 4.8/20.

Les méthodes proposées pour combiner des évaluations qualitatives et quantitatives sont clairement imparfaites, mais ces limites sont une conséquence directe des incohérences pédagogiques du système de notation quantitatif sur 20. En contrepartie, l'introduction de notes qualitatives dans un système quantitatif ne peut qu'améliorer la pertinence et la justesse de l'ensemble du processus d'évaluation. L'imperfection des méthodes de mise en œuvre de la période transitoire ne doit donc pas être un argument pour rejeter l'introduction progressive d'une évaluation qualitative et renoncer à améliorer notre système d'évaluation des acquis.

4 Conclusions

Ce livrable propose un système d'évaluation qualitative répondant aux objectifs recherchés. Il décrit une opérationnalisation de ce système à travers un système de notation permettant une mise en œuvre progressive.

Plusieurs arguments sont donnés pour affirmer que ce système est plus pertinent que la notation quantitative sur 20 pour attribuer une note représentative, qu'il permet de mettre en œuvre conjointement une évaluation formative et sommative, et enfin qu'il est au moins aussi précis pour effectuer des classements au sein d'une promotion.

La pertinence de la détermination d'un niveau de réussite permet de penser que le diplôme ainsi délivré sera plus lisible du point de vue de l'employabilité et pour des mobilités internationales.

Un travail reste naturellement à conduire pour expliquer cette démarche au sein du système universitaire français.

15. Cette amplitude est de 8 pour F, de 2 pour E et de $\frac{10}{3}$ pour chacune des trois autres notes.