



HAL
open science

Evaluation subjective appliquée à la notation d'étudiants

Christophe Simon, Alexandre Voisin, Lionel Grosjean, Eric Levrat

► **To cite this version:**

Christophe Simon, Alexandre Voisin, Lionel Grosjean, Eric Levrat. Evaluation subjective appliquée à la notation d'étudiants. CETSIS 2005, Oct 2005, Nancy, France. 6 p. hal-00172115

HAL Id: hal-00172115

<https://hal.science/hal-00172115>

Submitted on 14 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation subjective appliquée à la notation d'étudiants

C.SIMON¹, A.VOISIN², L. GROSJEAN¹, E.LEVRAT²

¹IUT d'Epinal, 7 rue des Fusillés de la résistance, 8810 Epinal,
{christophe.simon; lionel.grosjean}@univ-nancy2.fr

²CRAN, Faculté de Sciences, UHP Nancy 1, Bd des Aiguillettes,
54509 Vandoeuvre, {alexandre.voisin; eric.levrat}@cran.uhp-nancy.fr

RESUME L'évaluation d'un travail dans le cadre de jurys est un processus faisant appel à l'évaluation subjective. Nous proposons une méthode de synthèse d'évaluation d'avis d'experts à partir d'une interface web appliquée à l'agrégation de notes d'un jury de lecture de projets tuteurés. Nous explicitons le contexte qui a généré cette étude et la méthodologie de développement. Enfin, nous montrons l'application finale.

Mots clés : TICE, notation, évaluation subjective.

1 INTRODUCTION

Dans le cadre des formations de DUT, mais également dans le plupart des formations supérieures françaises, nous utilisons le projet tuteuré comme moyen pour donner de l'autonomie à l'étudiant. En DUT Organisation et Génie de la Production, nous proposons un sujet par année de formation. Le sujet est généralement, mais pas obligatoirement, proposé par un enseignant qui en définit la thématique.

A l'issue du projet, le travail de l'étudiant est évalué par un jury composé en partie de spécialistes de la thématique. La difficulté est alors de tenir compte des avis de chaque membre en restant si possible objectif. En outre, tous les membres du jury ne sont pas compétents sur tous les points d'évaluation et/ou ne savent pas (n'osent pas) défendre leur point de vue sur le travail de l'étudiant.

Pour tenir compte de ces facteurs et du caractère subjectif de la notation, nous en avons analysé le processus d'évaluation/notation et nous avons proposé l'application d'une théorie issue de la recherche à ce problème d'évaluation subjective.

2 ANALYSE

Le processus d'évaluation et notamment celui réalisé dans le cadre de jurys comporte des phases subjectives. De ce constat, nous avons proposé de réviser ce processus de notation afin d'en renforcer l'objectivité. Le projet tuteuré de première année convient parfaitement à cette problématique et nous a servi de support d'expérimentation. Ce projet est pris en charge par un ou plusieurs étudiants sur une durée prévisionnelle de 120 heures. Le jury utilise essentiellement le rapport final pour l'évaluation du travail réalisé. En outre, le jury est constitué de l'enseignant encadrant et d'autres enseignants relevant de disciplines diverses. Les membres du jury coordonnent leur avis durant une phase dite d'harmonisation. Cette phase peut être subjective car elle est un mélange d'explicitation de points de vue et de négociations.

Pour donner une meilleure visibilité à cette phase, nous avons défini les objectifs attendus, comme dans beaucoup d'autres formations. Ainsi, nous avons développé, à l'attention de l'étudiant, une fiche projet qui contraint le tuteur à une formalisation claire et précise ses attentes (Fig.1) et qui facilite ainsi la démarche d'évaluation du résultat. Cette grille a été mise en place sur notre intranet pour permettre à chaque intervenant de déposer ou de consulter la liste des projets tuteurés depuis n'importe quel poste informatique connecté. Cette solution est essentielle pour l'ensemble de l'équipe pédagogique puisque nous sommes sur un site délocalisé et l'équipe n'exerce pas exclusivement à l'IUT.



Fig.1 : Grille de proposition de projets

Nous avons également développé une grille de notation (Fig.2), mise à la disposition des membres du jury et des étudiants, qui précise les points spécifiques d'évaluation. Cette grille pose 3 grands critères décomposés en sous éléments. Ces 3 critères doivent mesurer les 3 exigences demandées à l'étudiant :

1. La compétence technique par la satisfaction des objectifs.
2. La conduite du projet.
3. La qualité de la communication écrite.

NOM Prénom: Année: 2004-2005	NOTE GLOBALE:	sur 20
GRILLE d'EVALUATION des PROJETS TUTEURS 1ère année (à compléter pour chaque étudiant par chaque membre du jury de lecture)		
<ul style="list-style-type: none"> • CONTENU TECHNIQUE : sur 7 		
1- Détermination du sujet: <i>Justification et analyse du contexte du sujet</i>	/1	
2- Maîtrise du sujet <i>outils et techniques utilisés</i>	/4	
3 - Conclusion	/2	
NOTE sur 7: _____		
<ul style="list-style-type: none"> • DEMARCHE STRUCTUREE DE CONDUITE DU PROJET : sur 6 		
1 - Planning prévisionnel d'avancement du projet	/2	
2 - Rencontres avec le tuteur	/2	
3 - Sources d'informations variées	/1	
4 - Gestion des aléas	/1	
NOTE sur 6: _____		
<ul style="list-style-type: none"> • PRESENTATION ECRITE : sur 7 		
1 - Structure (introduction, chapitres, conclusion)	/3	
2 - Orthographe - syntaxe	/2	
3 - Présentation - mise en page du texte:	/1	
4 - Clarté et qualité des figures, schémas, tableaux:	/1	
NOTE sur 7: _____		
Remarques éventuelles - Proposition d'un sujet de 2 ^e année s'il y a possibilité:		
IUT EFINSAL Département OGP		

Fig.2 : Grille initiale de notation

Si la pratique de cette grille de notation nous a permis d'éclairer l'étudiant et les membres du jury de lecture sur l'évaluation, elle n'a pas complètement satisfait à cet objectif.

La phase de jury reste une phase délicate dont le déroulement est lié aux domaines de compétences des membres du jury, à l'importance que les membres donnent à chaque critère et au déroulement de la séance d'harmonisation.

Pour donner un cadre de travail clair à cette séance, nous avons proposé l'utilisation des méthodes d'évaluation subjective et de fusion d'avis d'experts que nous avons développées et expérimentées dans le cadre de la recherche et du transfert de technologies. En outre, la solution proposée permet à chaque évaluateur de faire sa propre notation à distance grâce aux solutions NTIC.

3 LE SYSTEME D'ÉVALUATION

3.1 Le Questionnaire

L'objectif n'est pas de modifier les questionnaires établis mais de montrer comment la prise en compte d'une incertitude dans la notation peut être utilisée pour faciliter l'obtention d'une note finale.

Chaque partie du questionnaire fait référence à des compétences différentes, en technique, en conduite du projet et en communication écrite. Chaque enseignant, en qualité d'expert, dispose de compétences et un référentiel de valeurs propres. En cela, il est plus ou moins certain de son évaluation de chacune des parties. Pour agréger les évaluations, nous devons tenir compte de cette certitude. Elle est évaluée par l'enseignant lui-même sur une échelle comportant trois niveaux : expert, intermédiaire et néophyte.

3.2 Echelle de notation et évaluation du travail

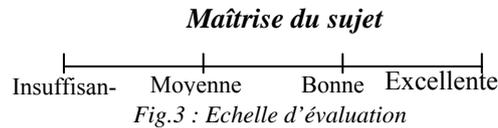
L'évaluation du travail est actuellement faite sur la base d'une valeur affectée par chaque évaluateur à chaque critère d'évaluation sur une échelle numérique. Dans notre démarche, nous remettons en cause ce principe de notation pour plusieurs raisons :

- une échelle numérique chiffrée peut entraîner une perte de dynamique dans la notation à cause de la conséquence associée à la valeur chiffrée,

- une note unique traduit rarement l'évaluation d'un travail, surtout pour un projet. Souvent l'évaluateur situe le travail dans un intervalle,
- une note unique ne permet pas lors de la phase d'agrégation de trouver un consensus autre que la moyenne. En proposant des intervalles, on peut éventuellement trouver une intersection permettant d'établir un consensus.

3.2.1 Echelle de notation

Plutôt que d'évaluer le travail du candidat sur une échelle numérique qui peut entraîner une distorsion notamment vers les valeurs extrêmes, nous proposons d'utiliser un axe continu délimité par des qualificatifs antagonistes aux extrêmes. La figure 3 en est un exemple.



3.2.2 Evaluation du travail

Pour recueillir les évaluations sous forme d'intervalles, nous les modélisons avec des distributions de possibilités. L'intérêt par rapport à une modélisation sous forme d'intervalles stricts est d'introduire une gradation dans les intervalles choisis par l'évaluateur.

Une distribution de possibilités est basée sur la notion d'ensemble flou proposée par Zadeh (1965). Un ensemble flou \bar{A} de \mathfrak{R} , où \mathfrak{R} est l'ensemble des réels, est défini par une fonction d'appartenance qui associe à chaque élément x du support un degré $\mu_{\bar{A}}(x)$ compris entre 0 et 1. Ce degré d'appartenance peut ensuite être interprété de différentes manières suivant l'utilisation que l'on souhaite en faire (Dubois, 1997). Une de ces interprétations concerne la notion d'incertitude pour laquelle l'ensemble flou représente une distribution de possibilités.

Une distribution de possibilités π de \mathfrak{R} est définie par un sous-ensemble flou de \mathfrak{R} de fonction d'appartenance $\mu_{\bar{A}}$ et satisfaisant une condition de normalisation :

$$\sup_x (\mu_{\bar{A}}(x)) = 1 \quad (1)$$

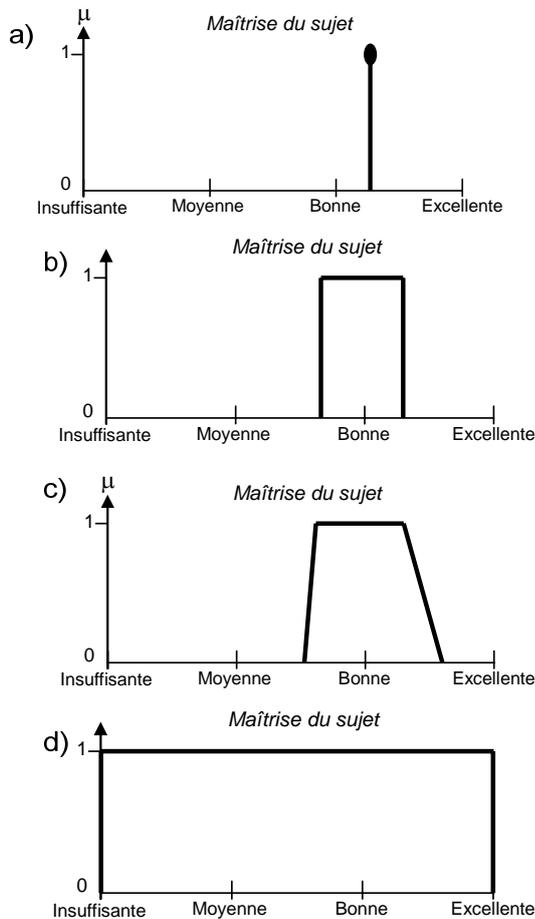


Fig.4 : différentes formes d'évaluation

$\mu_{\bar{A}}(x)$ représente le degré de possibilités qu'une variable X prenne la valeur x sachant uniquement que "X est \bar{A} ". Les valeurs du support de \bar{A} , $\{x / \mu_{\bar{A}}(x) \neq 0\}$, sont alors mutuellement exclusives, *i.e.* la variable X ne peut prendre qu'une valeur unique comme en probabilités. Le degré de possibilités ordonne les valeurs de \mathfrak{R} en termes de plausibilité.

Une distribution de possibilités peut être vue sous deux angles. Une vision verticale lorsque l'on considère les degrés d'appartenance $\pi(x) = \mu_{\bar{A}}(x)$ des éléments de \mathfrak{R} . Dans ce cas, π est vue comme la borne supérieure d'une famille de distributions de probabilités (Dubois, 1992). Une vision horizontale lorsque l'on regarde les α -coupes de \bar{A} , *i.e.* les sous-ensembles du support tels que $\mu_{\bar{A}}(x) \geq \alpha$ avec α dans $]0,1]$. Les α -coupes sont des ensembles classiques qui peuvent être vus comme un ensemble d'intervalles de confiance de niveau $1-\alpha$ (Dubois, 1993).

Pour simplifier la démarche d'évaluation, chaque évaluateur donne deux intervalles imbriqués. L'intervalle le plus large correspond au sous-ensemble de l'axe d'évaluation dans lequel l'évaluateur pense que l'évaluation se situe. Il correspond au support de la distribution de possibilités noté $S(\pi)$. L'intervalle le plus étroit correspond au sous-ensemble de l'axe d'évaluation dans lequel l'évaluateur pense que l'évaluation à le plus de chance de se trouver, il correspond au noyau de la distribution de possibilités, *i.e.* l' α -coupe de niveau 1, noté $N(\pi)$.

Cette démarche d'évaluation permet à l'évaluateur de traduire la connaissance qu'il a construite sur l'évaluation. La figure 4 montre différentes formes de distributions traduisant des connaissances différentes. La figure 4a présente le cas où l'évaluateur pense que l'évaluation n'a pas d'incertitude, *i.e.* $S(\pi)$ et $N(\pi)$ se réduisent à une valeur unique. La figure 4b représente le cas où l'évaluateur pense que l'évaluation est dans un intervalle pour lequel il est certain des bornes, *i.e.* $S(\pi)$ et $N(\pi)$ deux intervalles identiques. La figure 4c traduit le cas où l'évaluateur pense que l'évaluation est dans un intervalle dont il ne connaît pas les limites avec certitude, *i.e.* $S(\pi)$ contient $N(\pi)$. Enfin, la figure 4d caractérise le cas où l'évaluateur n'a pas d'avis sur l'évaluation, *i.e.* $S(\pi)$ et $N(\pi)$ sont identiques et égaux à l'ensemble de l'axe.

4 AGREGATION DES EVALUATIONS

Nous disposons pour chaque évaluateur e_i d'un ensemble de 11 évaluations, sous forme de distributions de possibilités π_{e_i, c_j} où c_j représente les critères classés en trois domaines. Nous avons également pour chaque évaluateur son niveau d'expertise $NE(e_i, d_k)$ dans chacun des trois domaines d'évaluation d_k .

Dans le cadre possibiliste, plusieurs modes d'agrégation des évaluations sont possibles. Le premier est le mode conjonctif. Il est utilisé lorsque l'on considère tous les experts comme fiables et correspond à l'intersection des évaluations. Ce mode est sensible à des évaluations discordantes. Lorsque l'on considère que dans un groupe d'évaluateur l'un d'eux est fiable mais sans savoir lequel, on utilise alors le mode disjonctif, *i.e.* l'union des évaluations. Ce mode peut conduire à des résultats peu informatifs. D'autres modes, que nous ne détaillerons pas ici, sont encore possibles.

Dans notre exemple, nous cherchons à agréger les évaluations en tenant compte des niveaux d'expertise de chaque évaluateur relativement aux trois domaines. Ainsi l'agrégation des évaluations va être réalisée pour chaque item de chaque domaine relativement à l'expertise des enseignants dans chacun des domaines. Le niveau d'expertise $w_{eik} = NE(e_i, d_k)$ d'un évaluateur vient modifier la distribution de possibilités représentant son évaluation π suivant la relation 2 (Sandri, 1995).

$$\pi'_{e_i, c_j} = \max(\pi_{e_i, c_j}, 1 - w_{iek}) \quad (2)$$

Ainsi, si l'expert est fiable, $w_{eik}=1$, et son évaluation n'est pas modifiée. Lorsque l'expert est moins fiable, $0 \leq w_{eik} < 1$, un niveau d'incertitude vient modifier l'évaluation. Finalement, si l'évaluateur n'est pas fiable du tout, $w_{eik}=0$, alors on obtient une distribution de possibilités égale à 1 sur l'ensemble du support, ce qui est équivalent à la figure 4d. Cette distribution représente l'ignorance sur la position de l'évaluation, puisque l'expert n'est pas fiable.

Dans notre exemple, nous avons trois niveaux d'expertise pour lesquels nous attribuons les valeurs *expert* $\rightarrow w_{eik}=1$, à intermédiaire $\rightarrow w_{eik}=0.6$ et à néophyte $\rightarrow w_{eik}=0.3$.

Finalement, les distributions résultantes sont agrégées suivant une règle conjonctive en les considérant comme fiables (Sandri, 1995) :

$$\pi'_{c_j} = \min_{e_i} (\pi'_{e_i, c_j}) \quad (3)$$

Dans le cas d'évaluations conflictuelles des experts les plus fiables, il peut arriver que le résultat ne soit pas normalisé, *i.e.* qu'une valeur au moins soit à 1. Dans ce cas, nous appliquons la règle de normalisation :

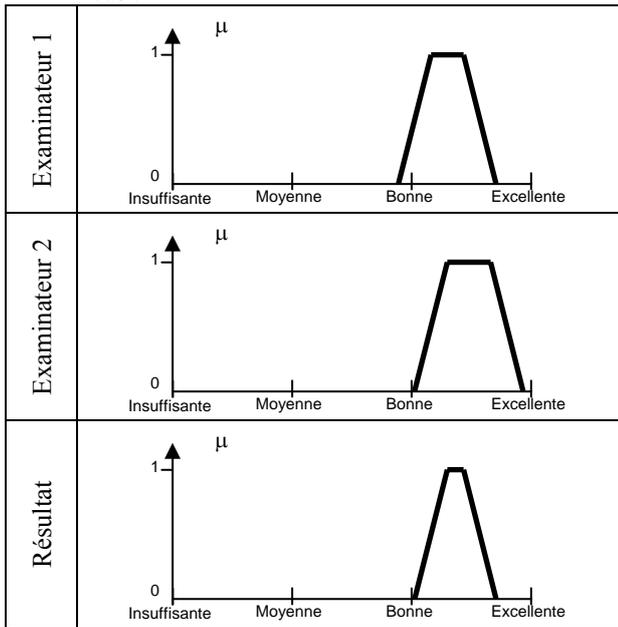
$$\pi'_{c_j} = \pi_{c_j} + (1 - \sup(\pi_{c_j})) \quad (4)$$

4.1 Exemples

Les exemples qui suivent montrent comment sont agrégées les notations de la compétence « *Maîtrise du Sujet* » de deux examinateurs dans des cas de figures différents.

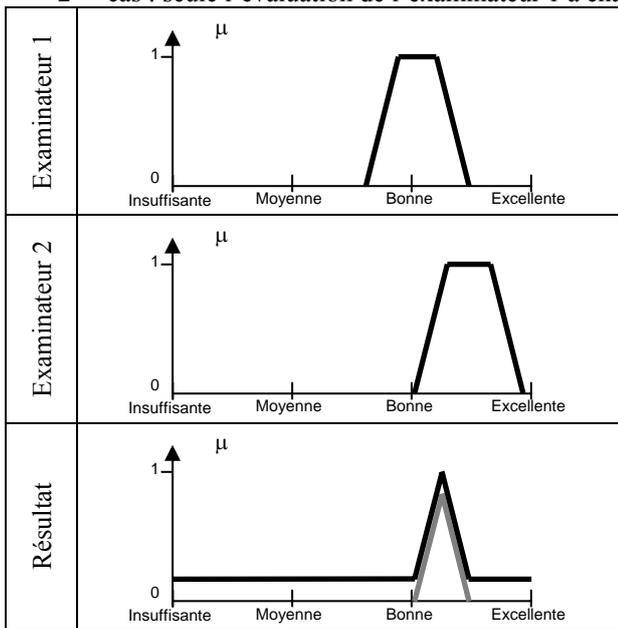
Dans ces trois premiers exemples, nous considérons que les deux examinateurs ont des niveaux d'expertise *expert*.

- 1^{er} cas :



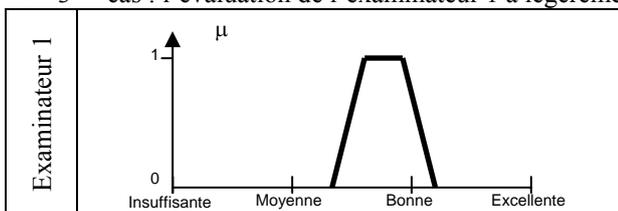
Dans ce cas, nous pouvons voir que les deux examinateurs sont d'accord, puisque les noyaux de leur évaluation ont une partie commune que l'on retrouve comme noyau du résultat.

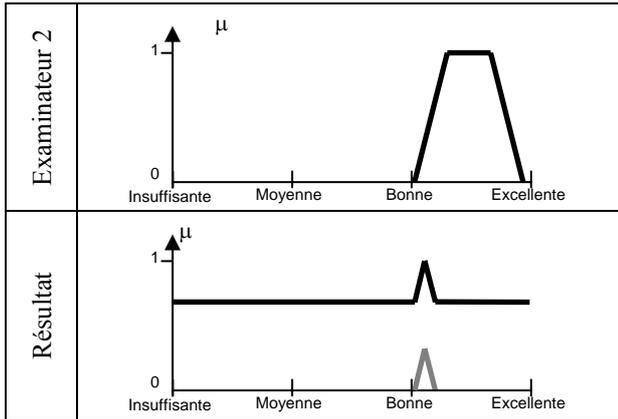
- 2^{ème} cas : seule l'évaluation de l'examinateur 1 a changée, elle a légèrement diminuée



Ici, les deux examinateurs ne sont pas tout à fait d'accord. Ainsi, les noyaux de leurs évaluations n'ont plus de zone commune. En gris clair on voit le résultat de l'agrégation qui n'est pas normalisé montrant ainsi le conflit. Le résultat final en noir après renormalisation, montre l'incertitude due au conflit existant entre les deux examinateurs. On peut voir que l'incertitude n'est pas très importante puisque le « pic » de l'évaluation se démarque nettement du « fond ».

- 3^{ème} cas : l'évaluation de l'examinateur 1 a légèrement diminuée

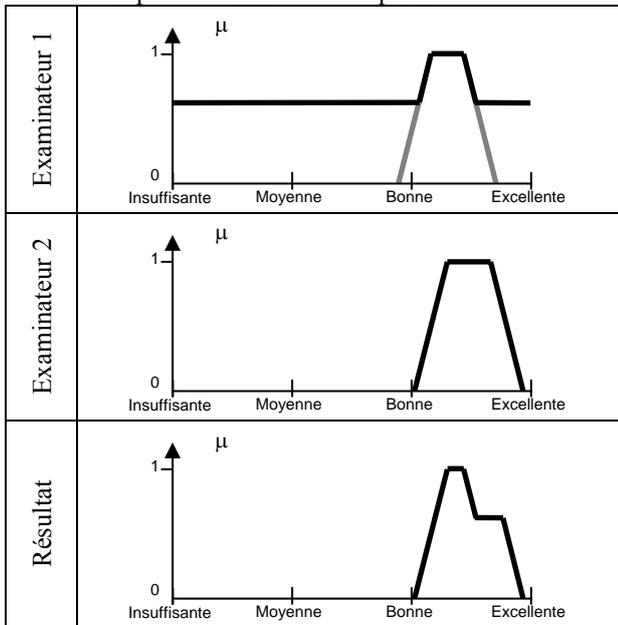




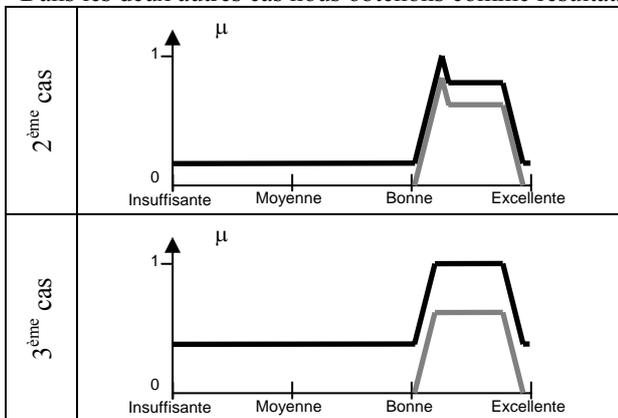
Ici, le conflit entre les examinateurs est beaucoup plus important, et au final, le « pic » de l'évaluation se démarque peu du « fond ».

Nous reprenons les trois exemples précédent, mais cette fois les examinateurs n'ont pas la même expertise : Le premier est néophyte tandis que le deuxième est expert.

- 1^{er} cas : pour l'examinateur 1, nous retrouvons en gris son évaluation, et en noir la distribution utilisée après prise en compte de son niveau d'expertise



Dans les deux autres cas nous obtenons comme résultats :



5 OBTENTION D'UNE NOTE GLOBALE

5.1 Utilité des évaluations

Pour chaque critère c_j , on définit une fonction d'utilité leur permettant d'être représenté dans un espace de notation. La fonction d'utilité va de l'espace d'évaluation, défini par l'axe d'évaluation, dans un intervalle de \mathfrak{R} . Dans notre cas, nous gardons pour chaque critère l'intervalle défini par le questionnaire initial (Fig.2).

La forme de la fonction d'utilité peut permettre de traduire différents comportements vis-à-vis de la notation. Par exemple, pour un travail moyen, un comportement sévère attribuera 8/20 et tandis qu'un comportement indulgent attribuera 12/20.

Dans notre cas, nous appliquons la même fonction à tous les critères sans préférence particulière. Aussi, nous utilisons une fonction d'utilité linéaire. La fonction d'utilité peut se représenter comme sur la

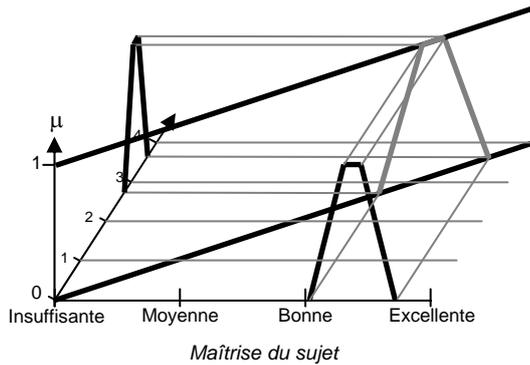


Fig.5 : Fonction d'utilité entre évaluation et note

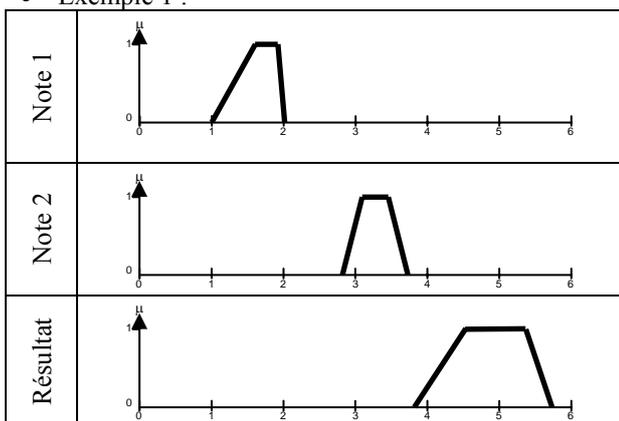
5.2 Agrégation des notations des critères

Dans le questionnaire initial, la synthèse pour la notation globale consiste à faire l'addition des notations de chacun des critères, sachant que la somme des notations maximales est égale à 20. Maintenant, pour chacun des critères nous avons une distribution de possibilités sur l'intervalle de notation du critère. Cette distribution de possibilités peut être considérée comme un nombre flou (Dubois, 1987). Nous utiliserons donc l'arithmétique floue (Dubois, 1987) pour additionner les notations des critères et obtenir la note globale.

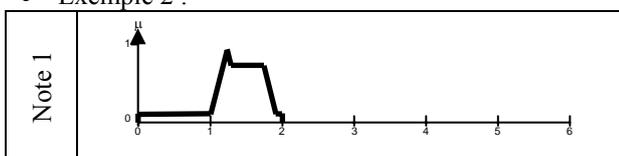
5.3 Exemples d'agrégation

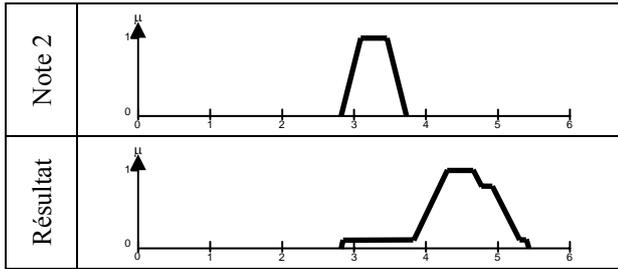
Nous montrons deux exemples d'agrégation de 2 notes, la première sur un référentiel de 2 points et la deuxième sur un référentiel de 4 points et donc le résultat sur un référentiel de 6 points.

• Exemple 1 :



• Exemple 2 :





5.4 Prise de décision

Le nombre flou obtenu par addition des critères doit donner une note unique. Nous proposons d'utiliser deux méthodes pour aboutir à ce résultat :

- Présenter le nombre flou aux enseignants lors du jury, pour servir de base à la discussion,
- Utiliser une méthode automatique pour obtenir une valeur à partir du nombre flou. Dans ce cas, nous proposons d'utiliser la moyenne des maximums.

6 APPLICATION

Le système d'évaluation subjective a été développé sur notre intranet pour permettre à chaque enseignant de coter les travaux de projets tuteurés à son rythme, quand bon lui semble, sans perturbation extérieure et surtout depuis n'importe quel poste informatique connecté à l'intranet.

La figure 6 montre l'interface web que nous avons développée dans l'environnement PHP/MySQL, permettant à chaque évaluateur de coter le travail. Les cotations sont stockées dans la base de données et le calcul de la note finale peut être refait à la demande. En outre, les étudiants peuvent consulter les résultats de l'évaluation ainsi que les commentaires transmis par les évaluateurs.

7 CONCLUSION

Ce système est actuellement en cours d'évaluation. Nous testons la validité de la démarche et surtout l'adhésion réelle des évaluateurs par rapport aux interfaces et au mode de cotation. S'il y a un a priori positif à la démarche, nous ne savons pas si la technique de cotation satisfera pleinement tous les acteurs du système.

REFERENCES

1. D. Dubois, H. Prade, The three semantics of fuzzy sets, *Fuzzy sets and systems*, 90, 1997, 141-150.
2. D. Dubois, H. Prade, S. Sandri, On possibility/probability transformation, in R. Lowen & M. Roubens (Eds), *Fuzzy Logic*, Kluwer Academic Press, 1993, 103-112.
3. D. Dubois, H. Prade, When upper probabilities are possibility measures, *Fuzzy sets and systems*, 49, 1992, 64-74.
4. D. Dubois, H. Prade (1987) Fuzzy numbers: an overview, in : James C. BEZDEK (Ed.), *Analysis of Fuzzy Information*, Boca Raton : CRC Press, 3-39
5. Sandri, S.A., Dubois, D., Kalfsbeek, H., Elicitation, Assessment, and Pooling of Expert Judgements Using Possibility Theory, *IEEE Trans. on Fuzzy Systems*, 3, 1995, 313-335.
6. L. A. Zadeh, Fuzzy Sets, *Information and control*, 8, 1965, 338-353.



Fig.6 : Grille de cotation subjective