Guide pratique de l'évaluation (littérale)

Jean Vaucher

Professeur titulaire
Groupe Incognito
Département d'informatique
et de recherche opérationnelle
Université de Montréal

octobre 1990
(révisé septembre 1991)

Avant-propos

Ce guide a été écrit dans le but d'aider mes collègues à attribuer des notes "correctes" avec facilité et confiance.

Tout d'abord, je devrais souligner que je n'ai aucune qualification ou diplôme particulier qui me permette d'imposer quoi que ce soit en matière d'évaluation. Par contre, cette année,[1] comme tous mes collègues, j'ai vécu l'introduction de la notation littérale - sans "mode d'emploi"- et j'ai dû réviser mes techniques d'évaluation, inventer un sens aux lettres de la notation, trouver des *trucs* pour calculer mes notes et refaire mes programmes d'ordinateur. Bref, passer beaucoup de temps à faire marcher la technique. Avec le *guide*, j'espère vous éviter mon travail de réinvention de roue.

Suite à des discussions sur le sujet, j'ai aussi remarqué que je n'étais pas le seul à avoir eu des problèmes à appliquer et interpréter la méthode littérale. En particulier, il ne semble pas y avoir consensus sur le sens des "lettres". Ceci est grave car si nous devons changer de système afin de nous *normaliser*, il faudrait s'entendre sur cette *norme* afin de pouvoir s'y conformer. De la situation actuelle, où chacun est forcé d'inventer sa méthode, se dégage une impression de flou et d'arbitraire qui pourrait facilement mener à la contestation de la légitimité des notes.

Pour éclaircir la situation, j'ai contacté des "sources officielles" à l'Université et j'ai posé la question suivante: «quelle devrait être la moyenne d'une classe moyenne au premier cycle?» Dans mon esprit, ça aurait pu être "C" avec 2 niveaux plus haut (A et B) et 2 niveaux plus bas (D et échec)... Dans une deuxième étape de réflexion, constatant qu'un étudiant doit maintenir une moyenne de "C" pour réussir un programme de premier cycle, j'ai déduit que si chaque professeur fixait sa moyenne à "C", la moitié de nos étudiants échouerait sur la moyenne générale. Quelle moyenne donc faudrait-il viser afin que le système marche: C+, B-, B...? La réponse "officielle" ne m'a pas aidé. «Il n'y a pas de norme, m'a-t-on dit, ça dépend des départements et des titulaires.»

Suite à cette expérience, j'ai décidé qu'il serait utile de passer quelque temps à mettre sur papier ce que mes collègues et moi avions appris, conclu ou élaboré cette année en matière d'évaluation.

Je dois avouer que je n'ai pas fait de recherche bibliographique en bonne et due forme. J'ai bien ouvert Bloom[2] au départ, mais je l'ai vite refermé et ce n'est qu'à la toute fin du travail que je suis tombé sur les monographies de notre Service Pédagogique. Deux volumes sont pertinents et j'en recommande la lecture. Le premier[3] souligne très bien les problèmes de l'évaluation; le second[4] donne de bons conseils sur la composition d'examens à choix multiple. Cependant, ni l'un ni l'autre ne traite vraiment des problèmes étudiés ici. De plus, ces documents ont le

---

[1] 1989-1990

[2] Bloom, Benjamin, *Taxonomy of educational objectives*, 1956.

[3] *Dossier sur l'évaluation*, présenté par France Lafontaine, Service Pédagogique, Université de Montréal, 1979 (9,00$).

[4] Bernard, Huguette et F. Fontaine, *Les questions à choix multiples*, Service Pédagogique, Université de Montréal, 1979 (11,00$).

désavantage d'être des documents "institutionnels". En quelque sorte, ils engagent l'Université et la réputation académique des auteurs. Le style se doit d'être tempéré afin de ne pas susciter la contestation. Le présent "guide" est un document personnel qui n'engage pas l'Université. Si jamais des contestations étaient basées sur mon texte, il serait facile de désavouer l'amateur que je suis. Je peux donc être aussi catégorique et pragmatique que je juge nécessaire et le lecteur sera libre de ne retenir que ce qui lui est utile.

Par rapport à la version précédente d'octobre 1990, ce texte comporte quelques retouches à la section sur l'échelle "B" (p. 19) et à la section sur la normalisation des notes (pp. 21-22).

Remerciements

Par leurs commentaires, suggestions et corrections, plusieurs personnes ont contribué à améliorer ce guide. Je tiens à remercier D.B., F.B., G.B., P.B., C.C., G. L. , L.L., B.L., U.M., C. S-V. et P.M.

Introduction

Les pédagogues distinguent entre la mesure et l'évaluation.[5] La mesure, c'est un score obtenu dans un test quelconque (p.e. 15 sur 20); l'évaluation, c'est l'interprétation de ce score (p.e. moyen ou excellent). Notre ancien système de notation, basé sur les pourcentages, est axé sur les tests et il s'apparente à la mesure, tandis que la notation littérale s'oriente plus vers l'évaluation. L'avantage de la notation littérale dans les bulletins est de pouvoir faire ressortir le message essentiel d'une évaluation: est-ce que l'étudiant a réussi ou échoué, est-il supérieur à la moyenne, mérite-t-il une bourse, etc...

Pour que le processus d'évaluation fonctionne correctement, il faut qu'il y ait entente générale sur les différents jugements qu'il est possible d'émettre et sur la manière dont on peut les exprimer par le biais d'une seule lettre. De plus, il faut trouver un système facile à appliquer et à justifier afin de limiter l'arbitraire (ou l'impression d'arbitraire) pour que tous les intervenants, c'est-à-dire les professeurs, les étudiants et les lecteurs des bulletins, aient confiance dans l'exactitude du processus.

Le simple fait d'utiliser la notation littérale dans les bulletins officiels ne résout pas le problème fondamental de l'évaluation: comment combiner et traduire des résultats de tests en évaluation. Et ceci, avec cohérence, facilité et certitude.

Dans ce qui suit, je présenterai d'abord les extraits du règlement pédagogique qui décrivent la notation littérale ainsi que l'échelle de conversion de la Commission des études. Ensuite, je décrirai certains problèmes typiques rencontrés avec le nouveau système, puis je proposerai un schéma d'évaluation qui procède en deux étapes: distinguant en premier lieu, la réussite de l'échec, pour ensuite attribuer des mentions (lettres) appropriées. Finalement, j'exposerai certaines techniques que j'ai trouvé utiles et je montrerai comment elles s'appliquent dans des cas concrets tirés de mon expérience.

[5] F. Fontaine, «Evaluer? Mesurer?» dans *Dossier sur l'évaluation*, Service Pédagogique, Université de Montréal, 1979.

Définition du système de notation littérale[6]

Le règlement pédagogique de la Faculté des Arts et Sciences (section 7.1) décrit le système de notation littérale par la table suivante:

Lettres	Points	Mention
A+	4,3	
A	4	Excellent
A-	3,7	
B+	3,3	
B	3	Très bien
B-	2,7	
C+	2,3	
C	2	Bien
C-	1,7	
D+	1,3	
D	1,0	Passable
E	0,5	Echec (faible)
F	0	Echec (nul)

Certains seuils ajoutent un “sens” à ces lettres. Au premier cycle, la note de passage pour un cours est D ou mieux. De plus, on exige une moyenne générale de 2,0 (C) ou mieux pour la réussite d’un programme. Ces deux seuils correspondent à des moyennes de 50% et 60% selon l’ancien système. Il y a un troisième point de repère dans nos règlements, c’est la moyenne minimale requise pour être admissible à la maîtrise: 2,7 ou “B-” dans un système et 70 % dans l'autre.

Aux études supérieures, les exigences sont plus élevées. La note de passage pour un cours est "C" (article 33 du règlement pédagogique de la FES) et la moyenne générale à maintenir est "B-" ou 2.7 (articles 42 et 56).

Finalement, il existe une table de conversion entre pourcentages et lettres qui a été adoptée par la Commission des études. Cette table a été largement diffusée par les étudiants, et on la retrouve dans certains agendas. Nous la reproduisons ci-dessous. L’aspect détaillé de cette table suggère que son emploi est une garantie de “l’objectivité” de l’évaluation. **C’est loin d’être le cas! et je critiquerai plus loin l'emploi abusif qui peut en être fait!** Effectivement, cette table a été adoptée uniquement pour fins de conversion, au 31 août 1989, de la moyenne cumulative antérieure des bulletins.

[6]Un tableau des systèmes de notation en vigueur au Québec est donné à l'annexe 3.

A+	90-100	
A	85-89	Excellent (80-100)
A-	80-84	
B+	77-79	
B	73-76	Très bien (70-79)
B-	70-72	
C+	65-69	
C	60-64	Bien (57-69)
C-	57-59	
D+	54-56	
D	50-53	Passable (50-56)
E	35-49	Échec (faible)
F	0-34	Échec (nul)

Table de conversion pour la moyenne cumulative des bulletins

Problèmes de la notation littérale

Un premier problème vient de l'absence de normalisation des termes (et lettres). Pour que la notation littérale puisse faire ressortir le message essentiel d'une évaluation, il faut savoir quelle lettre dénote quel message. Selon notre règlement, voici les "messages" associés à l'évaluation.

A	Excellent
B	Très bon
C	Bon
D	Passable
E	Échec (faible)
F	Échec (nul)

C'est bien d'avoir une liste de mots; c'est mieux de s'entendre sur le sens de ces mots. Sur des dossiers français, on peut trouver des mentions différentes: "assez bien", "bien" et "très bien". Une université américaine utilise les termes: "conditional pass", "satisfactory", "good" et "very good". Comment comparer notre "très bon" avec un "assez bien" ou un "good". De même, est-ce que le "A" est attribué à 5% des étudiants, 10%, ...? Il est illusoire de penser obtenir un standard absolu, mais il serait souhaitable que certains critères soient énoncés tout haut afin de réduire les divergences d'interprétation.

Un deuxième problème découle de l'application aveugle de l'échelle de la Commission des études pour convertir en lettres des résultats d'examens exprimés en pourcentages. Comme nous l'avons déjà dit, un pourcentage dénote une *mesure* tandis qu'une lettre implique une *évaluation*. L'application d'une échelle présuppose des résultats "normalisés"; on ne devrait pas s'en servir directement

avec des notes brutes.[7] Par contre, une échelle peut servir d'étalon pour vérifier (a posteriori) la cohérence d'un barème de correction.

Je m'explique par un exemple. Disons que dans un examen donné, mes deux meilleurs étudiants obtiennent 67% et 69% respectivement. Si j'applique sans réfléchir l'échelle de conversion, je dois leur attribuer la note C+. En fait le raisonnement devrait se faire à l'inverse: typiquement, les meilleurs étudiants dans une classe normale décrochent des A (entre 80 et 100% selon l'échelle). A défaut d'autres informations qui expliquent ce résultat, je dois conclure qu'il y a un problème: soit dans mon examen, soit dans ma correction et la réaction doit être de corriger le problème, (par exemple, en redressant les notes).[8]

On n'a pas toujours besoin de convertir de pourcentage à lettre. Certaines formes de travaux se prêtent naturellement à l'attribution directe de mentions: examens oraux, dissertations, projets et travaux pratiques. Dans ces cas, on peut procéder avec le modèle de "piles". Initialement on fait une première distinction entre les bons, les moyens et les échecs (les piles). Puis, on procède à un classement de chaque pile en sous-piles par comparaison. On peut associer les lettres directement aux piles, utilisant les variantes "+" et "-" s'il y a vraiment des différences remarquables entre sous-groupes. Avec ces formes de travaux, la pertinence et l'utilité de la notation littérale sautent aux yeux.

Cependant, on n'a pas contourné pour autant le problème de conversion de notes car, il faut souvent combiner et pondérer les résultats de plusieurs épreuves pour obtenir une note globale. Comment faire? Ceci nous amène à un troisième problème car **l'utilisation des "points" associés aux lettres (A=4, B=3, etc...) s'est avérée spécialement non-recommandable** pour la combinaison de notes littérales. Un exemple illustre le problème.

Prenons un cours comportant 4 épreuves avec pondérations égales. Il est clair qu'après avoir bien réussi la première épreuve avec "A" (4 points), l'étudiant peut s'abstenir de se présenter aux autres épreuves, car un calcul de la moyenne selon les "points": (4+0+0+0)/4 donne 1.0 ou "D" et il est assuré de la note de passage.

Un peu de réflexion montre **l'erreur** dans ce raisonnement. L'usage des points par l'Université pour le calcul de la moyenne cumulative va de pair avec la règle que chacun des cours doit être réussi (avec D [1.0] ou mieux). Un cours échoué ne contribue pas à la moyenne; mais il doit être repris ou remplacé. Appliquant cette philosophie aux épreuves à l'intérieur d'un cours, on conclut que l'échec à une épreuve entraîne l'échec au cours. Donc, loin de réussir, notre étudiant coule misérablement! Si l'exigence de réussite de chaque épreuve n'est pas ce que vous cherchez, il faut faire autre chose. C'est là qu'une échelle de conversion comme celle de la Commission peut être utile. Soit A=86% et 0% pour l'absence, nous avons

---

[7] L'usage institutionnel qui a été fait de l'échelle de conversion se justifie, dans le cas de moyennes cumulatives, parce que la loi des grands nombres atténue les différences individuelles de notation entre épreuves et normalise les résultats.

[8] Je ne prétend pas qu'on doit toujours ajuster les notes afin que les meilleurs obtiennent des "A". Le niveau des étudiants admis varie d'année en année et le hasard fait que certaines classes sont exeptionnellement faibles. Le sujet sera traité plus en détails dans une section sur la "normalisation" des notes.

une moyenne de (86+0+0+0)/4 = 21,5% (ou F) qui traduit bien la situation.[9] Facile... à qui sait comment s'y prendre!

COROLLAIRE: **ne pas laisser à d'autres** (votre secrétariat, la Faculté, le centre de calcul ou le bureau du Registraire) **le soin de combiner des évaluations littérales d'épreuves partielles**; il est probable qu'ils utiliseraient la "méthode des points" avec les aberrations qui en découlent. Il est recommandé[10] de faire vos propres calculs et de soumettre une seule note littérale globale.

Une échelle de conversion est donc utile pour la combinaison de notes littérales, mais toutes les échelles ne sont pas aussi bonnes les unes que les autres. En particulier, une échelle non-linéaire comme celle de la Commission peut donner des surprises . Par non-linéaire, j'entends que l' intervalle en pourcentiles n'est pas égal pour chaque mention. Par exemple, A+ recouvre un intervalle de 10% tandis que B- recouvre un intervalle de 3%. Selon l'ordre dans lequel on combine les notes, il peut y avoir des différences marquées. Par exemple, prenons deux notes avec pondérations égales: 100% (A+) et 72%(B-). Quelle est la mention globale? Première technique: moyenne de 86%, ce qui donne "A". Deuxième technique, passage par les "points": (4.3+2.7)/2 = 3.5, livrant "A-" ou "B+" selon l'arrondi. De telles anomalies ne peuvent qu'entraîner un sentiment d'arbitraire et d'injustice.[11] Il faudra faire attention à nos échelles!

Les grandes lignes de l'évaluation

Le premier objectif d'une évaluation est de différencier la réussite de l'échec. Cette décision est importante. Si on décrète la réussite, on se porte garant de la compétence du candidat et on engage sa réputation. D'un autre côté, le constat d'échec a des répercussions importantes. L'étudiant doit reprendre le cours. Il peut perdre un semestre, une année, ou pire. C'est aussi un gaspillage pour l'Université, car l'étudiant a monopolisé des ressources qui auraient mieux servi à d'autres fins. Comment trancher dans les cas limites? Dilemme!

En réalité, la situation n'est pas si difficile à traiter, car le système prévoit une zone grise entre la réussite et l'échec. Voici les trois niveaux qu'il faut départager:

- Niveau de compétence[12] exprimé par les mentions A, B ou C.[13]
  --- seuil à 60% ---
- Niveau d'incertitude (réussite conditionnelle, passable), mention D.
  --- seuil à 50% ---
- Constat d`échec, mentions E ou F.

---

[9] L'utilisation de poids négatifs (E= -1 et F= -2) aurait aussi le même effet.

[10] Même obligatoire me souligne un collègue.

[11] Le paragraphe est inspiré d'un incident vécu au Certificat en Droit: une étudiante avec 29,5 sur 30 à l'intra et 47 sur 70 au final (pour une moyenne pondérée de 76.5%) pensait se voir attribuer au moins un B ou un B+ selon l'échelle; elle reçut "B-". Justification: conversion de chaque épreuve en lettre et utilisation des points pour la moyenne.

[12] Le terme "compétence" n'est pas consacré; mais il semble singulièrement bien approprié à la situation.

[13] A moins d'avis contraire, la mention d'une lettre comme C comprend aussi ses sous catégories (C+,C et C-).

Les seuils de l'ancien règlement pédagogique expriment assez bien ces trois notions: échec, passage minimal et compétence. Le seuil de 50% traduit le fait que l'étudiant devrait connaître au moins la moitié de la matière de chaque cours et se trouver plus près de la connaissance parfaite que de l'ignorance totale. Le 60% reflète le fait qu'un diplôme doit attester d'un niveau de maîtrise additionnelle.

La notion de compétence recouvre toute une gamme d'habiletés. On sait que dans tout domaine, certains praticiens sont meilleurs que d'autres et que la majorité sont moyens (voire très moyens); mais chaque praticien compétent doit pouvoir fonctionner utilement et il doit pouvoir se démarquer des personnes qui n'ont pas reçu sa formation. Par analogie, prenons des métiers: il y toutes sortes de plombiers, mais on s'attend à ce que chacun puisse installer un robinet, souder des tuyaux et déboucher une toilette. De même, tout journaliste doit pouvoir écrire un texte lisible.

De manière pratique et mesurable dans le contexte universitaire, **l'attestation de compétence dans un domaine (mentions A,B ou C) signifie que l'étudiant est prêt à aborder d'autres activités (emplois ou cours) qui découlent du domaine en question.** Si votre évaluation ne vous permet pas d'attester ce degré de compétence, il faut l'indiquer avec D, E ou F.

La mention D (passable) dénote une situation où l'étudiant a démontré une connaissance minimale des notions importantes du cours sans faire preuve de maîtrise ni de facilité avec la matière. On ne voit pas l'utilité à faire reprendre le cours, mais l'étudiant n'est pas mûr pour aller plus loin.

A l'UdeM, un D représente une faiblesse qui peut être rachetée par des résultats supérieurs ailleurs afin de maintenir une moyenne de C. A McGill, le sens du D est légèrement différent, plus proche des critères opérationnels énoncés plus haut:

> «Même si le 'D' est une note de passage, il ne donne pas accès aux cours dont il constitue une des conditions préalables, ni n'est reconnu si le cours est obligatoire dans le programme de l'étudiant.» [14]

Traitons maintenant de l'échec (mentions E ou F). Dans chaque matière, il y a un niveau minimal de connaissances que tout étudiant devrait acquérir. **Si un niveau minimal de connaissances n'est pas atteint, nous devons constater l'échec**.

Comme l'échec a des conséquences sérieuses sur le déroulement des études d'un étudiant, il faut être absolument sûr de ce jugement. Un examen bien conçu peut aider la décision. Imaginons un examen avec des sections "faciles", d'autres "moyennes" et "difficiles". Vous pourrez conclure à l'échec, non seulement parce qu'un seuil (comme 50%) n'a pas été acquis mais aussi par la nature des erreurs... Par exemple, "question X représente le B-A-BA du cours et le candidat n'a même pas pu y répondre!!!" De façon pratique, ceci veut dire qu'un collègue pourrait facilement corroborer votre conclusion. Si vous n'êtes pas sûr de votre jugement d'échec.... Utilisez la zone grise, le D.

---

[14] Explication des barèmes, Bureau du Registraire, McGill.

Autre conseil: avec un seuil à 50%, laisser des notes entre 49 et 50% invite à la contestation.... Je les augmenterais à 50% (D). Par contre, je n'irais pas plus loin dans cette voie. Déjà, à 49% l'étudiant est loin de la "compétence" qu'on souhaite. N'oublions pas le rôle des jurys de fin de session pour les cas limites. C'est au jury à examiner globalement le dossier de l'étudiant et à tenir compte de circonstances particulières afin de redresser des anomalies grossières.

Compléments sur l'échec

Pourquoi est-ce que les étudiants échouent? Est-ce parce qu'ils sont fondamentalement mauvais et ne devraient pas être à l'université? Pas nécessairement. Il est vrai que certains n'ont pas les aptitudes requises pour certains cours. D'autres travaillent à temps partiel et n'ont pas investi l'effort requis pour réussir. Des étrangers ont le mal du pays. Les uns n'aiment pas le prof ou la matière, les autres sont déprimés. Etc... etc... Ce ne sont pas les résultats d'examens qui vont nous donner la bonne réponse!

J'ai dit que l'échec indique qu'un niveau minimal de connaissances n'a pas été atteint; j'ajoute ici que **c'est tout ce que ça indique**. Un échec ne permet pas de conclure qu'un étudiant est voué à l'échec éternel dans une matière donnée et encore moins de conclure qu'un étudiant est "mauvais" globalement. Cette façon objective d'envisager la chose devrait permettre à l'étudiant (comme au professeur) de voir l'échec de façon moins péjorative et plus constructive.

Doit-on prévoir un certain taux d'échec dans un cours ? Question traître! Faites très attention au sens que vous attribuez au mot "prévoir". Doit-on faire couler du monde? Est-ce qu'un taux d'échec élevé est la garantie d'un cours sérieux... ou la marque d'un mauvais enseignant? L'idéal serait de ne pas avoir d'échecs. **Mais, on ne planifie pas l'échec; on le constate**. C'est à dire qu'on ne peut pas fixer, a priori, un taux d'échec.

Cependant quiconque a enseigné -surtout en première année- sait qu'il y a un nombre impressionnant de candidats qui ne réussissent pas. Mon modèle personnel de la situation est que certains étudiants sont fascinés par le mystère des disciplines pour lesquelles ils ont le moins de talent. Dans ce cas, le meilleur service qu'on peut leur rendre, c'est de leur faire prendre conscience qu'ils sont dans la mauvaise voie. Et ça le plus vite possible. En quelque sorte, je vois les examens de première année comme des séances d'orientation.

Dans mes cours de première année, je constate une perte[15] d'étudiants de l'ordre de 20 à 30% et j'estime que la première année fait partie des tests d'admission. Par contre, si ce "triage" est nécessaire en première année, il n'a plus sa place après.[16] Ceux qui pensent assurer la qualité par l'échec devraient consulter Bernard[17] et les

---

[15]J'utilise le mot "perte" car il est difficile de distinguer entre échec et abandon volontaire. Par perte, j'entends des étudiants dont on a eu trace car ils ont écrit un examen ou soumis un travail mais qui n'ont pas réussi le cours.

[16]Dûs à des circonstances fortuites ou personnelles, l'échec reste possible; mais un taux d'échecs plus élevé que 5% devrait être considéré anormal.

[17]Huguette Bernard, «Les systèmes de notation», dans *Dossier sur l'évaluation*, Service Pédagogique, Université de Montréal, 1979, pp. 101-103.

références qu'elle cite sur le manque de corrélation entre les notes et le succès professionnel.

Classification de la réussite

Une fois que l' épineux problème des échecs a été réglé, le reste est plus facile.

> La population des étudiants qui ont réussi est relativement homogène et on peut attribuer des mentions selon la proportion des étudiants qu'on s'attend à avoir à chaque niveau.

A mon avis, les meilleurs 50% devraient recevoir des A et B tandis que les autres 50% se verraient attribuer les mentions C et D. J'ai maintenant répondu à la question que j'avais posée aux instances officielles: «quelle doit être la moyenne d'une classe moyenne? » On peut déduire que la note médiane pour les étudiants qui ont réussi doit se situer à mi-chemin entre C et B.

En plus de détails, voici une répartition qui semble raisonnable avec des mentions descriptives:

A	20%	- mérite une bourse (*magna cum laude* et peut-être admissible au PhD???)
B	30%	- admissible à la maîtrise (*cum laude*)
C	40%	- compétent
D	10%	- passable

Il y a deux points importants à souligner. Premièrement, la répartition ne touche que ceux qui ont réussi et les proportions suggérées ne sont pas influencées par le taux d'échecs. L'échec est décrété par un mécanisme différent (compétence minimale) où les proportions n'interviennent pas. Deuxièmement, les proportions visent le long terme; il est clair que la proportion exacte des "A" va varier d'une année à l'autre selon la force des promotions.

Pour terminer, voyons A+, les *summa* cum laude. Devrais-je suggérer un quota comme 1% de la population? Pas nécessaire d'aller si loin. Chaque année on trouve des étudiants qui nous rendent heureux d'être professeurs. Non seulement ils obtiennent les meilleures notes, leurs questions en classe démontrent qu'ils anticipent ce qu'on va dire. Dans leurs examens, on trouve des éléments de réponse qui dépassent ce qu'on a enseigné. Souvent, ils ont quelque chose à nous apprendre et on aimerait qu'ils poussent les études jusqu'au doctorat afin qu'ils puissent un jour prendre notre place. Ce sont eux les A+.

Élaboration d'une échelle de conversion

Une échelle de conversion donne l'équivalence entre des notes exprimées en pourcentage et des mentions littérales. Par exemple, l'échelle de la Commission indique que la mention "très bien" ou B correspond à une note entre 70 et 79%. D'une certaine façon, une échelle associe un jugement de valeur à une mesure du degré de connaissance d'une matière. Par exemple, selon l'échelle de la Commission, la réussite d'un cours exige la connaissance d'au moins la moitié de la

matière (D=50% ou mieux); de même, la connaissance de 80% (ou plus) de la matière représente un résultat excellent.

Le premier rôle d'une échelle de conversion est de servir de modèle explicatif et opératoire de l'évaluation. Pour les étudiants, une échelle est un gage d'objectivité. Elle permet de juger le degré de réussite et de prévoir la note finale.

Deuxièmement, une échelle peut servir à la combinaison d'évaluations littérales car l'échelle traduit les évaluations en "degrés de connaissance" qui eux se combinent très bien mathématiquement.

Troisièmement, dans la mesure où une note en pourcentage est le **reflet exact** du degré de connaissance d'une matière, une échelle peut servir à la conversion de pourcentages en lettres. Mais, attention! Si un examen est mal conçu, ce n'est pas l'emploi aveugle d'une échelle de conversion qui produira une **évaluation** correcte ou objective.[18] En fait, la philosophie de la notation littérale préconise la normalisation des résultats numériques avant l'application d'une échelle de conversion - j'en reparlerai plus loin.

Malgré ma critique de l'utilisation irréfléchie des échelles d'équivalences, je reste convaincu qu'il est essentiel pour chaque professeur d'avoir son échelle de conversion et de la communiquer aux étudiants. Par contre, cette échelle se doit d'être différente de celle de la Commission des études.

Dans ce qui suit, nous traitons d'abord les contraintes que les échelles doivent respecter, puis les problèmes que ces contraintes entraînent. Ensuite, nous regardons des échelles utilisées ailleurs et nous examinons des distributions de notes obtenues à l'Université de Montréal pour établir l'applicabilité de diverses propositions. Finalement, je proposerai deux échelles pratiques.

Prenons d'abord les contraintes. Le règlement pédagogique définit trois seuils de réussite et ceci détermine implicitement des équivalences. Premièrement, les mentions E et F dénotent l'échec, tandis que D (ou mieux) indique la réussite dans un cours; ceci correspond à l'ancien seuil de 50%. Ensuite, une moyenne de 2,0 (C) est nécessaire pour la réussite d'un programme (ancien palier à 60%) et 2,7 (ou mieux) est nécessaire pour l'admission aux cycles supérieurs (70% selon l'ancien règlement).

Quand on veut aller plus loin, il y a des problèmes. Premièrement, la notation décimale nous a habitué à situer nos seuils et nos notes à des multiples de 5% ou 10%, mais le nombre de mentions qu'il faut accommoder (soit 4 ou 11 selon notre règlement pédagogique[19]) ne divise pas élégamment les 50% qui représentent la plage de réussite. Avec ce nombre de mentions, doit-on accepter des intervalles bizarres (p.e. 3.57%) ? Avoir des intervalles de tailles variables? Changer le nombre de mentions utilisées?

Autre considération dans le choix des intervalles: les intervalles choisis devraient répartir les étudiants en groupes de manière utile et significative.

---

[18]Pour éviter le cumul des problèmes de conversion, on peut tout noter numériquement, comme avant, et ne convertir en lettre que la note globale.

[19] Les 4 mentions principales (A,B,C et D) ou 11 mentions si on compte les sous-divisions: A+,A,A-,B+,B,B-,C+,C,C-,D+ et D....(sans D-).

Pour avoir une idée d'alternatives possibles, je vais présenter la succession des échelles en usage général à McGill depuis 1937, - pas nécessairement parce que McGill fait mieux les choses que d'autres mais parce que McGill utilise la notation littérale depuis assez longtemps et parce que ces échelles sont disponibles (publiées par le Registraire). Un tableau des autres systèmes de notation en vigueur au Québec est donné à l'annexe 3.

McGill (1937 - 1971)

%	Mention	
80 - 100	A	(1° classe)
65 - 79	B	(2° classe)
50 - 64	C	(3° classe)
0 - 49	F	échec

McGill (1971 - 1980)

%	1° cycle	2°/3° cycles	Pondération
80 - 100	A	A	4.0
65 - 79	B	B	3.0
55 - 64	C	F	2.0
45 - 54	D		1.0
0 - 44	F		0.0

McGill (1980 - ... présent)

%	1° cycle	2°/3° cycles	Pondération
85 - 100	A	A	4,0
80 - 84	A-	A-	3,7
75 - 79	B+	B+	3,3
70 - 74	B	B	3,0
65 - 69	B-	B-	2,7
60 - 64	C+	F	2,3
55 - 59	C		2,0
50 - 54	D		1,0
0 - 49	F		0,0

On remarque certaines constantes dans ces échelles. Dans chacune, les mentions A et B correspondent aux mêmes intervalles de notes: 80-100 et 65-79 respectivement. Il y a moins de paliers que prévus dans notre règlement mais les intervalles sont plus réguliers (multiples de 5%). Il y a deux différences principales avec l'échelle de la Commission: d'abord, la limite entre les C et les B se situe à 65% et pas à 70% comme chez nous. Ensuite, le nombre de mentions entre B et F est très réduit à McGill: la zone grise, notre D,D+ et C-, est représentée par une seule lettre, le D. Finalement, exception faite du A+ qui n'existe pas à McGill, notons que les pondérations associées aux lettres (A=4.0, A-=3,7, etc...) sont identiques chez eux comme chez nous.

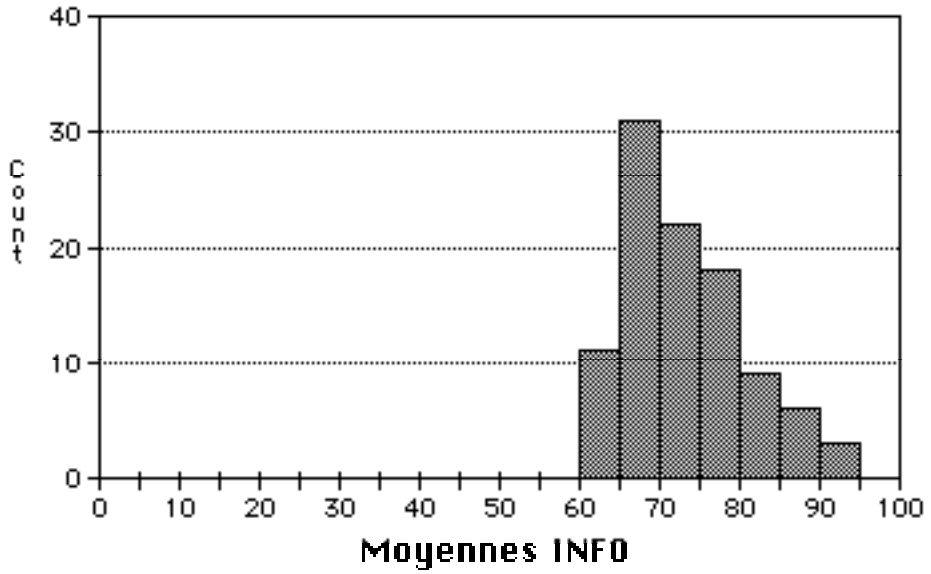
Les échelles ci-dessus ont l'avantage d'être plus simples que celle de la Commission; mais, est-ce qu'elles reflètent nos habitudes en matière d'évaluation? En particulier, pour fixer le seuil entre B et C (65% à McGill), il faudrait savoir quelle est la note médiane de nos étudiants. Je cherchais aussi un seuil raisonnable pour départager les A des B.

Pour répondre à ces questions, j'ai fait une analyse rapide des bulletins des étudiants dans mon Département.

Regard sur les moyennes cumulatives

Plus précisément , j'ai fait l'analyse des moyennes cumulatives en date de mai 1989 (avant la conversion au système littéral) pour divers programmes du Département d'informatique et RO.

Le premier graphique montre la répartition des moyennes de 100 étudiants choisis au hasard dans nos programmes spécialisé, majeur et mineur**.** Ces étudiants avaient réussis entre 10 et 84 crédits. Les chiffres sont biaisés par le fait que les échecs n'entrent pas dans le calcul de moyenne et que les étudiants avec moyennes inférieures à 60% sont éliminés; les moyennes rapportées ici sont donc plus élevées que les moyennes obtenues. Néanmoins, ceci nous donne une idée de la répartition des notes des étudiants qui réussissent.



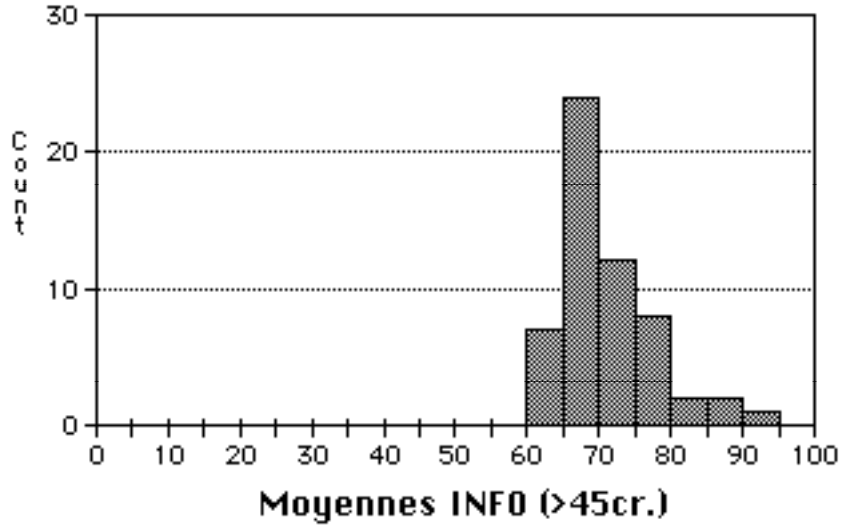
moyenne: 72.9
médiane: 71.3

résultats > 80%: 18%
résultats > 85%: 9%
résultats > 90%: 3%
maximum: 91.8

Selon ces données, un seuil à 70% entre C et B et un seuil à 80% entre B et A correspondraient à peu près à nos objectifs: d'accorder des A et B à environ 50% des nos étudiants et des A à 20%.

Résultats d'étudiants "avancés":

Si on limite l'échantillon aux 56 étudiants "avancés" qui ont réussi 45 crédits ou plus, nous avons des moyennes encore plus centrées autour de 70%:



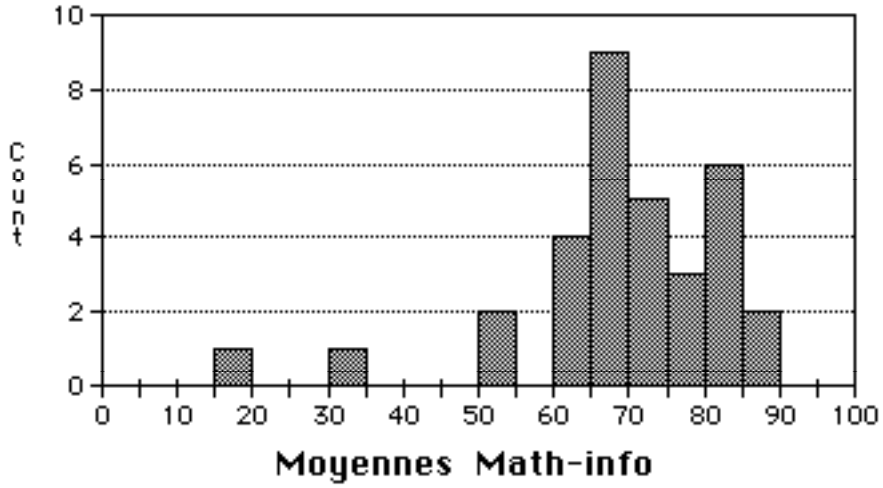
moyenne: 70.8
médiane: 69.0

résultats > 80%: 9% (5 étudiants)
résultats > 85%: 5% (3 étudiants)
résultats > 90%: 2% (1 étudiant)
maximum: 91.8

Ici aussi, 70% correspond bien à une mesure de milieu.

Résultats du programme MATH-INFO:

Voici les résultats pour tous les étudiants du programme (33 personnes y compris 2 échecs):



moyenne: 68.6
médiane: 68.3

résultats > 80%: 24% (8 étudiants)
résultats > 85%: 6% (2 étudiants)
résultats > 90%: 0%
maximum: 87.4

De l'ensemble de ces résultats, je conclus que 70% semble être un palier raisonnable (en informatique et RO) pour départager la moitié des étudiants en dessus de la “moyenne” (A et B) de ceux en dessous de la “moyenne” (C et D). De même, utiliser 80% pour délimiter les B des A semble utile. Environ 10 à 20% de nos diplômés obtiendraient un “A” pour le BSc.

Attribution de notes: échelle “A”

Basé sur toutes ces données, je propose une première échelle qui est linéaire, simple à comprendre et à appliquer et qui récompense nos meilleurs étudiants (les A+).

Mention	Plage	Note représentative[20]
A+:	90 - 100%	95%
A:	80 - 90 %	85%
B:	70 - 80 %	75%
C:	60 - 70 %	65%
D:	50 - 60 %	55%
E:	40 - 50 %	45%
F:	0 - 40 %	0%

Je recommande cette échelle pour l’attribution directe de notes littérales.

Il y a quelques années, un de mes collègues, le professeur Paul Bratley, proposait un barème de correction pour nos projets de fin d’études (le cours IFT3051). Ce barème est presque identique à l’échelle pratique “A”. La description du niveau de compétence associé à chaque catégorie est excellente et le barème a souvent été repris comme modèle. Je le reproduis ci-dessous en ajoutant entre parenthèse la mention pertinente (A+,A, etc...).

[20] La borne inférieure exacte devrait être choisie telle que les résultats (arrondis) affichés par des programmes de “spreadsheet” ou “tableurs” correspondent aux plages. C’est à dire que la borne inférieure exacte pour A+ serait 89.5%, pour A 79.5%, etc...

IFT3051 - Projets de fin d'année en informatique

Barème d'évaluation

A titre indicatif, je vous suggère le barème suivant pour vos projets. Vous êtes bien sur libre de l'ignorer ou de le changer: c'est vous qui accordez des notes à vos étudiants, pas moi. Je vous demande cependant de ne pas pousser votre générosité au point de manquer de toute discrimination.

90-100 La perfection (A+): le projet devrait normalement inclure une contribution originale de l'étudiant sur le plan des idées ou des méthodes. Un simple travail de programmation- même très bien exécuté- ne devrait pas normalement mériter un 100%.

80-90 Un excellent travail (A): bien testé, bien documenté, et bien conservé pour une utilisation subséquente. Si l'étudiant n'a pas combiné des idées originales, ou au moins il a accompli le travail avec flair et élégance.

70-80 Un bon travail compétent (B): Peut-être le projet n'a pas avancé autant que prévu, mais ce qui est fait est bien testé et bien documenté; peut-être le travail accompli est excellent, mais la documentation n'est pas tout à fait satisfaisante (elle devrait cependant être adéquate pour quelqu'un qui voudrait continuer le projet plus tard).

60-70 On commence à manifester un **certain mécontentement** (C). Peut-être un programme existe, mais il n'a pas été testé sérieusement; peut-être le travail n'est pas documenté de façon adéquate. Le projet pourrait difficilement être repris et continué par une autre personne.

50-60 La déception (D): L'étudiant a accompli assez de bon travail pour qu'on lui accorde ses crédits, mais le projet n'est pas utilisable: il y a trop d'erreurs ou de lacunes, ou tout simplement le projet n'est pas terminé à la date voulue.

< 50 L'échec (E et F): A votre avis, le travail de l'étudiant n'est pas suffisant, en quantité ou en qualité, pour qu'on lui accorde des crédits.

Je vous suggère aussi d'utiliser la note 49 pour des cas douteux: si l'ensemble du dossier le justifie, le jury pourra monter la note à 50. Si vous voulez que l'échec soit définitif, mettez une note en bas de 45.

Pour des projets "presque" terminés, je vous suggère d'accorder une note basée sur l'état réel du projet, quitte à changer avant les réunions du jury (mi-mai), plutôt que d'accorder une note plus généreuse basée sur des promesses.

Paul Bratley (1984)

Combinaison de notes: échelle “B”

Pour combiner les résultats de plusieurs évaluations littérales, il est utile d'avoir une échelle de conversion avec des gradations plus fines. On cherche aussi à avoir une échelle *linéaire* afin d'éviter les anomalies que nous avons relevé plus tôt[21]. Après plusieurs essais, voici une proposition qui couvre toutes les mentions de notre règlement (sauf D+) avec des notes représentatives (“milieu” d’intervalles) centrées sur des multiples de 5%.

Mention	Équivalence	Plage
A+	100%	92.5 - 100
A	90%	87.5 - 92.5
A-	85%	etc ...
B+	80%	
B	75%	
B-	70%	
C+	65%	
C	60%	
C-	55%	
D	50%	47.5 - ...
E	40%	29.5 - ...
F	0%	0 - 29.5

Avec cette échelle, les notes équivalentes pour C et D correspondent exactement aux seuils de 50% et 60% de l’ancien règlement. En contre-partie les bornes inférieures des intervalles ne tombent plus sur des déciles. Notez qu’on a étendu légèrement D vers le bas (de 50% à 47.5%) et que la limite entre B et A est passée de 80% à 82.5%.

Je recommande cette échelle pour combiner des résultats exprimés en notes littérales[22].

[21] En particulier, on évitera l'emploi des *points* (A=4,...D=1) ou de l'échelle non-linéaire de la commission des études

[22] Ne pas utiliser cette échelle pour convertir directement un **score** numérique en lettre avant d'avoir lu la section suivante sur la normalisation des notes.

La normalisation des notes

Il est très difficile de concevoir un bon examen: un examen qui mesure correctement les connaissances de l'étudiant selon les objectifs pédagogiques du cours et qui exprime cette mesure avec des notes qui cadrent avec nos échelles.

Dans une situation statique (matière figée et titulaire à long terme), on peut améliorer les examens d'année en année en vérifiant le degré de pertinence et de discrimination de diverses questions. Mais qui de nous se trouve dans une telle situation! En réalité, la matière évolue d'année en année et les professeurs changent de cours tous les trois ans. Plus un professeur tient son enseignement à coeur, plus il aura tendance à innover... et plus il lui sera difficile de concevoir de bons examens ! Il est donc certain que les notes seront biaisées: trop hautes, trop basses ou trop peu réparties! Que faire?

Une solution, c'est la normalisation des notes (le *tripotage* diront certains). Par normalisation, j'entends l'ajustement des notes brutes obtenues dans un examen suite à la détection de défectuosités dans cet examen[23]. Ceci dans le but de rendre les notes plus représentatives de la compétence réelle des étudiants.

Mais, *ATTENTION !* Toute manipulation réduit la crédibilité d'une évaluation et **il faut que les notes restent perçues comme un résultat direct du travail de l'étudiant** et non pas comme le résultat de manipulations obscures du titulaire. Il faut donc assortir le processus de certaines garanties pour les étudiants. Voici deux principes que j'ai pu dégager.

Principe 1: ne jamais réduire une note!

Donc, si jamais je donnais un examen "trop facile" où tous auraient entre 95 et 100%, les notes resteraient telles quelles. Pas question de réduire les notes pour avoir une moyenne plus *raisonnable* .

> [Corollaire: il vaut mieux concevoir un examen trop difficile que trop facile... vous pourrez toujours corriger ensuite .]

Principe 2: utiliser une formule d'ajustement simple.

Entre nous, je pense qu'il est illusoire d'essayer d'obtenir la fameuse "cloche" par une transformation quelconque. La "cloche" représente les grands nombres, à long terme. Nos examens, c'est plutôt les petits nombres à court terme.

De plus, l'évaluation parfaite est une chimère; toute mesure comporte une marge d'erreur et il en est de même avec nos examens. Ayant admis que les notes ne pourront jamais être exactes, l'objectif premier de la normalisation transparaît: c'est d'éviter les erreurs systématiques ou grossières.

Revenons à la pratique de la *normalisation*. La première étape, c'est la détection d'un problème; la deuxième (si nécessaire), c'est l'ajustement des notes.

---

[23] Dans un premier temps, je traite la normalisation *traditionnelle* de résultats numériques. Comme nous le verrons plus tard, la notation littérale permet de normaliser sans donner l'impression de tripotage.

Je prends pour acquis deux choses: 1) basé sur votre expérience, vous avez composé le meilleur examen possible et 2) vous avez une idée préalable de la distribution attendue des notes. Une différence marquée entre les résultats attendus et les résultats obtenus sera un indice de défectuosité dans l'examen. Remarquez que c'est vous qui devez décider de ce qui est *normal* et de ce qui ne l'est pas pour une situation donnée.

Comme exemple de *normalisation* - et sans prétention que ma méthode soit ni la meilleure ni même la bonne - je vais décrire comment je procède, faisant l'hypothèse d'une promotion moyenne avec sa part de bons étudiants.

Je commence avec une liste triée des notes ou un histogramme. Avec un examen valable, je m'attends à trouver 10% (ou plus) des notes en dessus de 80% et au moins une dans les 90%. Si c'est exact et si il n'y a pas un nombre exorbitant d'échecs (d'après mon expérience du cours), je juge l'examen valable et le processus ne va pas plus loin. Si les notes semblent trop fortes, je les laisse tel quelles et je prévois augmenter la difficulté de la prochaine épreuve. Mais, si les notes me semblent trop faibles, je vais chercher un facteur multiplicatif à appliquer (à chaque note) pour que mes meilleurs étudiants obtiennent des "A". Pour appliquer ces corrections à la main sans calculatrice, j'utilise une variante simplifiée de la multiplication: pour chaque intervalle de 5 ou 10 % je calcule et j'applique une correction fixe.

Prenons un exemple. Disons que la meilleure note obtenue est 82% et que je considère que ceci représente un "A+". Une façon simple et rapide pour *normaliser*, serait d'ajouter 9 points à toutes les notes entre 80 et 90% (le 82% deviendra un 91%), 8 points à celles entre 70 et 80,... et 1 point à celles entre 10 et 20%.

Il reste le cas où l'examen est trop discriminatoire. Les bons étudiants sont bien notés mais, pour une raison ou une autre, on a l'impression qu'il y a trop d'échecs. Ce cas est rare, mais je l'ai rencontré en première année pour des examens "objectifs" avec correction stricte. Si on peut décider d'un seuil de réussite, on peut toujours diviser les notes de *passage* en intervalles égaux correspondant aux mentions de l'échelle "A" ou "B" et attribuer soit les mentions soit les notes équivalentes.

Certains professeurs sont rebelles à l'idée d'ajuster les notes. Ils disent «si notre meilleur étudiant n'a répondu correctement qu'à 5 questions sur 10, son *score* est 5 sur 10 ou 50% et il est malhonnête de prétendre autrement. Comment donc justifier de remonter sa note à 80 ou 90%». C'est ici qu'il est utile de distinguer entre score et évaluation car même si le score est 5 sur 10, il est possible que ceci représente un excellent résultat qui devrait être noté 80 ou 90% selon nos barèmes types.

C'est ici que la notation littérale devient utile car on peut faire la normalisation dans le passage des scores aux lettres (en établissant un seuil de réussite et en divisant les notes de passages en intervalles égaux). La distinction est maintenant claire: le score est un nombre (objectif et immuable), tandis que l'évaluation est une lettre (obtenue du score avec les corrections qui s'imposent).

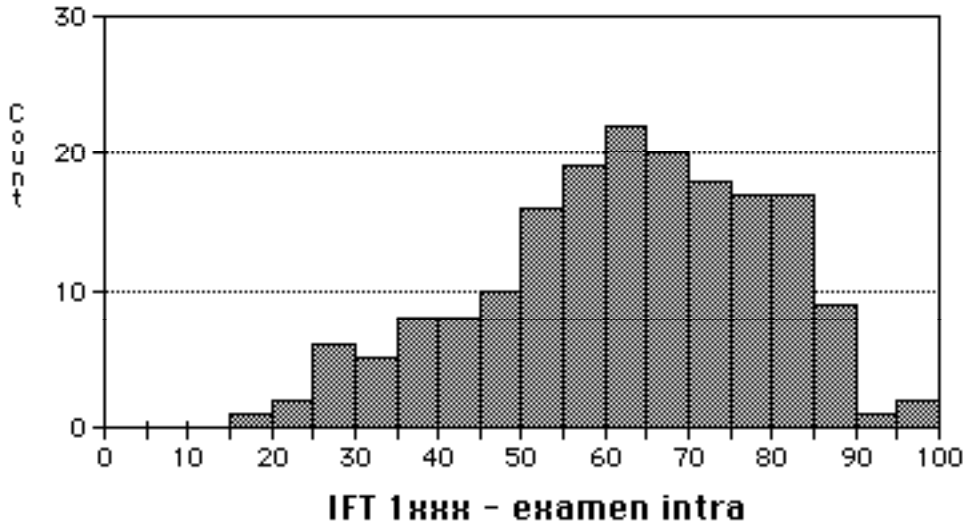
Cependant, les principes énoncés plus haut restent valables. Les examens devraient être conçus afin que les scores correspondent le plus possible aux évaluations normales (échelles "A" ou "B"). Un autre schéma de conversion ne sera

utilisé que pour éviter des erreurs systématiques ou grossières et cet autre schéma ne devra pas entrainer une réduction de note.

Exemples concrets

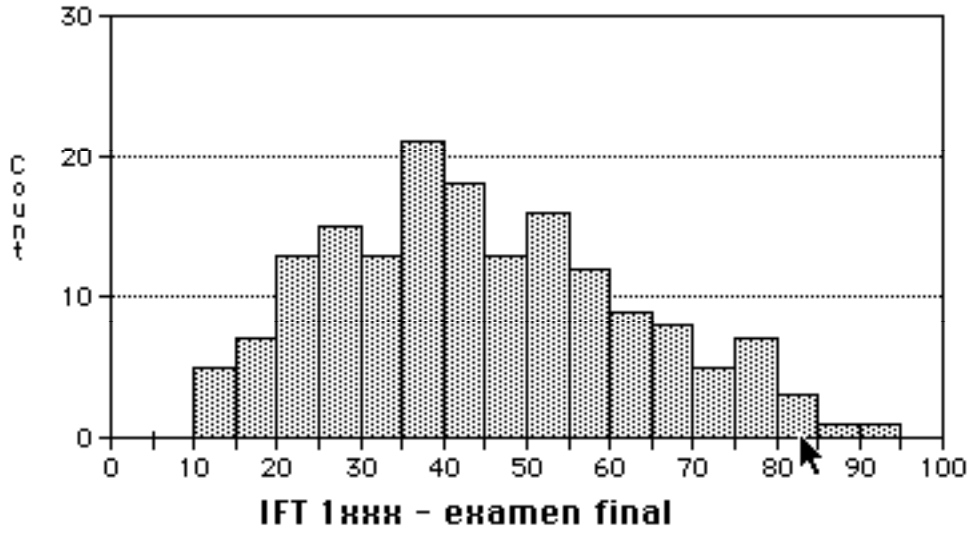
Prenons quelques exemples (camouflés pour protéger les innocents) pour voir comment les principes précédents pourraient s'appliquer.

CAS 1: examen intra de première année



Résultat typique de première année. L'examen semble valide: bonne répartition de notes, moyenne de 62% et médiane de 63% et 23% d'échecs. Pas d'ajustement nécessaire.

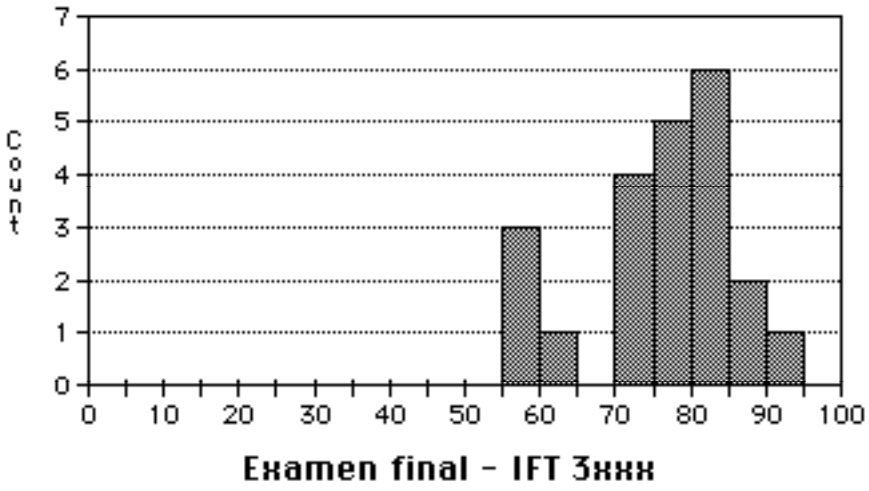
CAS 2: examen final de première année



Ici, l'examen semble nettement trop dur et discriminatoire: moyenne de 44%, médiane de 42%, une majorité d'échecs et juste 5 étudiants en dessus de 80%. De plus, c'est le même groupe que pour le cas 1 exception faite d'abandons (des plus faibles). On voit que les résultats ne concordent pas avec ceux de l'intra.

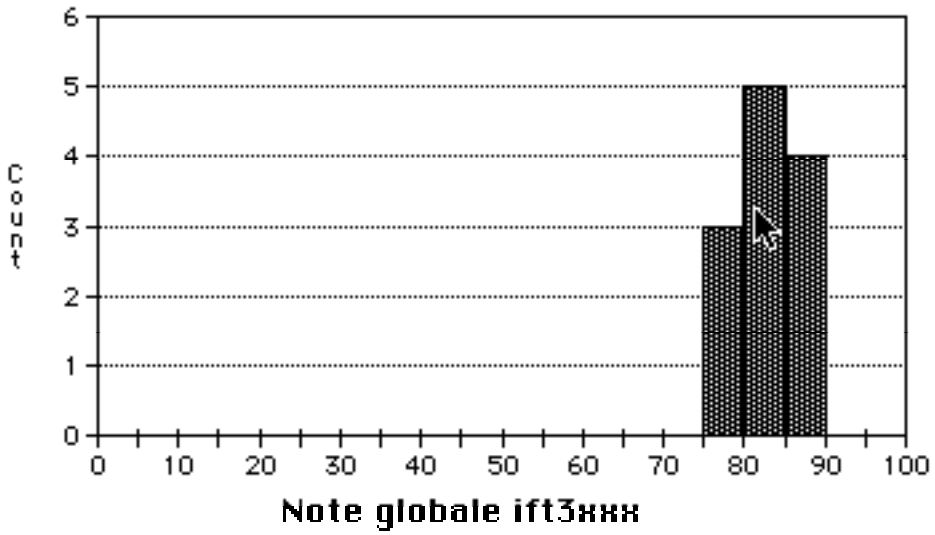
Solution: ajustement global. On fixe la note de réussite à 30% et on répartit les 10 paliers de l'échelle "B" (D,C-,C,C+,B-,B...A+) sur la plage de 30% à 100%.
... On essayera de faire un meilleur examen la prochaine fois !

CAS 3: examen final de dernière année



Examen discriminatoire. Résultats semblent valable. Aucune correction.

CAS 4: notes globales en dernière année



Ici on a un cours à option avec des étudiants motivés. Les bons résultats ne sont pas surprenants. L'écart est assez restreint. Ceci reflète l'homogénéité de la classe; et le fait que la note est le résultat combiné de 5 épreuves.

Autres techniques

Dans ce guide, j'ai abordé les principes et la pratique de l'évaluation dans un cadre assez général.

Reste que certains types d'épreuves amènent des problèmes particuliers. L'annexe 1 traite un tel cas: l'emploi de la notation littérale pour des examens objectifs avec un nombre limité de questions.

Il y a aussi des pratiques locales qui méritent d'être plus largement connues. L'annexe 2 décrit les barèmes avec seuils employés au Département d'informatique et recherche opérationnelle.

CONCLUSIONS

Au début, j'envisageais ce guide sous forme d'un feuillet explicatif de 4 ou 5 pages; mais, après réflexion, j'ai été amené à traiter le problème de façon plus globale et le texte s'est allongé. Pour résumer:

- Le guide souligne la différence entre la mesure et l'évaluation. Notre ancien système de notation, avec des pourcentages, s'apparentait à la mesure, tandis que la notation littérale veut afficher le message essentiel de l'évaluation.

- On note plusieurs problèmes avec notre pratique actuelle de la notation littérale: l'absence de normalisation, l'abus des échelles de conversion et l'emploi de techniques viciées pour la combinaison de notes.

- Je vous propose de faire l'évaluation en deux étapes:
  - d'abord, en vous basant sur le critère de "compétence", décidez s'il y a réussite (notée par A, B ou C) ou échec (noté avec E ou F) - la mention D (passable) indique que vous n'avez pas pu trancher.
  - en second lieu, pour les étudiants **qui ont réussi**, octroyez les mentions selon le succès relatif de telle sorte, qu'à long terme, la moitié reçoivent des A et des B - et l'autre moitié, des C et des D.

- Suite à l'examen de la pratique à l'Université de Montréal et ailleurs, le guide propose deux échelles opérationnelles: l'échelle "A" pour l'attribution directe de notes littérales et l'échelle "B" pour toute combinaison de lettres ou pour conversion entre lettre et pourcentage (normalisé).

- Un principe est clair: l'évaluation littérale implique la normalisation des résultats. Cependant, il faut que les notes restent perçues comme un résultat direct du travail de l'étudiant. Je dégage deux recommandations: utiliser une formule d'ajustement simple et ne jamais réduire une note!

Je ne m'attends pas à ce que vous soyez d'accord avec tout ce qui a été exposé ici. Toutefois, j'ose espérer que le guide contribuera à normaliser le processus d'évaluation et vous aidera à attribuer des notes justes avec facilité et avec confiance.

Annexe 1:
La correction des tests objectifs

Dans les sections précédentes, nous avons vu comment convertir en notation littérale des notes exprimées en pourcentage. C'est à dire comment passer d'un système avec un très grand nombre de résultats possibles (100 et *plus avec des décimales*) à un autre qui a un nombre plus restreint de possibilités: 7 mentions avec l'échelle "A", 12 avec l'échelle "B" ou 13 selon le règlement pédagogique.

Ici je veux traiter un problème différent: comment établir la bonne mention à partir d'un système avec un nombre limité de résultats, par exemple: un test composé d'une seule question objective avec deux résultats possibles: bon ou mauvais (PASS / FAIL). S'il semble approprié de mettre F pour un mauvais résultat, que donner pour la bonne réponse? C? A? A+?

Initialement, le problème m'a été rapporté par collègue sous une forme assez différente: «Je veux employer la notation littérale pour mes tests, me disait-il, mais je n'ai pas assez de questions !»

Après explications, j'ai compris ce qu'il voulait dire. Dans son cours, il donnait assez souvent des tests objectifs avec 5, 10 ou 15 questions. Maintenant, pour tenir compte de toutes les mentions, il lui faudrait au moins 22 questions pour distinguer tous les niveaux: 11 (A+, A, A-, jusqu'à D) pour couvrir les mentions de la plage de réussite (50 à 100%) et un nombre égal pour la plage des échecs (0 à 50%). Comment faire avec moins de questions?

Première réponse: ne pas utiliser la notation littérale! Continuer avec le vieux système de pourcentages pour les épreuves intermédiaires et ne faire la conversion en notation littérale qu'à la toute fin du cours pour la note globale.

Mais, si on voulait vraiment employer les lettres, que faire?

Le principe fondamental qui va nous guider, c'est que la précision de la notation doit correspondre à la précision de la mesure. En physique, on parle des chiffres significatifs; on ne doit pas exprimer la longueur d'un bout de bois comme étant 12,042 cm si l'erreur estimée est ±0,1 cm, on dira plutôt 12,0 cm.

Donc, pour une question VRAI/FAUX, les résultats sont parfait ou mauvais, 100% ou 0%, A+ ou F. Comme avec la correction numérique, c'est par le cumul de ces évaluations binaires qu'on obtient une évaluation plus raffinée .

Exemple:

> Dans une suite de 15 questions VRAI/FAUX, un étudiant obtient 11 bonnes réponses. Quelle est sa note globale?
>
> On convertit en pourcentage avec 100% pour une BONNE réponse et 0% pour une MAUVAISE. L'étudiant a donc une moyenne de 11*100% divisé par 15, soit 73%: un **B** selon l'échelle "A" ou l'échelle "B" .

Avec une bonne réponse en moins (dix), on obtiendrait, 100% * 10 / 15 = 66%: un **C** selon l'échelle "A" ou un **C+** selon l'échelle "B". Notez que, selon ce calcul, il est impossible d'avoir **B-**. Mieux vaut se limiter aux mentions A,B,C,D et F de l'échelle "A" et ne pas essayer d'attribuer des "+" et des "-".

Ma conclusion générale sur les tests objectifs est qu'il est inutile de donner des mentions plus fines que A,B,C,etc... tant qu'il y a moins de 25 points distincts dans le schéma de correction. On utilisera plutôt un barème de correction basé approximativement sur l'échelle "A". Le barème exact pourra être fixé *a posteriori* pour permettre la normalisation (revoir section pertinente).

Voici quelques exemples de barèmes pour des tests objectifs avec divers nombres de questions. Dans chaque cas, on associe D à la réussite de 50% (au moins) des questions et on suit (approximativement) l'échelle "A".

	Résultats obtenus				
Note littérale	6 questions	7 quest.	8 quest.	9 quest.	10 quest.
F	0,1	0,1,2	0 - 2	0 - 3	0 - 3
E	2	3	3	4	4
D	3	4	4	5	5
C	4	5	5	6	6,7
B	5	6	6	7	8
A	6	7	7,8	8,9	9,10

Avec moins de 6 résultats possibles, on utilisera encore moins de mentions. Par exemple, dans un de mes cours je donne une suite d'exercices à réaliser en laboratoire. Pour chaque exercice, il y a trois résultats possibles: Réalisé dans les temps, correct mais en retard (1 semaine maximum) et échec. Présentement, je donne 0,1 ou 2 points; si j'utilisais la notation littérale, je donnerais F,D ou A.

En conclusion, pour des épreuves avec un nombre limité de résultats possibles, notez en points ou utilisez un petit nombre de lettres basé sur les équivalences de l'échelle "A".

Annexe 2:
Les barèmes avec seuils

Je traite ici d'une technique utile qui semble particulière au Département d'informatique et RO. Typiquement, beaucoup de cours comprennent des travaux pratiques et des exercices qui servent à l'assimilation de la matière. Dans l'évaluation du cours, doit-on tenir compte de ces travaux ou est-ce que l'évaluation doit se faire surtout sur la base d'examens? Le problème avec les travaux pratiques, c'est qu'ils sont peu discriminatoires. Le but d'un travail pratique, c'est de faire apprendre; les travaux se font en équipe, il n'y a pas de limite de temps stricte, *ils sont conçus pour être réussis*: il est donc difficile d'y échouer. D'un certain coté, on voudrait donner une forte pondération à ces travaux pour refléter l'effort investi; mais, souvent on constate que les examens n'ont plus grand effet sur la note finale.... Il devient quasiment impossible de couler un cours! Pour les cours avancés ceci n'est pas trop grave, mais en première année, c'est un problème sérieux car il faut pouvoir **discriminer** afin de réorienter le plus rapidement possible (par l'échec) les étudiants mal placés.

La solution en vigueur au Département, c'est le "barème avec seuil." Par ceci, on entend que les travaux pratiques sont contributoires seulement si les examens sont réussis. Les seuils habituels étaient 50% ou 40%; en littéral, ça pourrait être D ou E. Si le seuil n'est pas atteint, la note de travaux pratiques est limitée à la note d'examen (ou au seuil).

Voici un exemple avec des pourcentages. Soit un barème avec 40% pour les examens et 60% pour les TPs et "seuil" à 50%. Disons que l'étudiant a 80% dans les TPs et 45% à l'examen (seuil non franchi). Sans seuil, le calcul pondéré donnerait (80 x 0.6 + 45 X 0.4) = 66% ou réussite comme résultat final. Avec le seuil, la note finale reste à 45% (échec).

Annexe 3:
Tableau comparatif des systèmes de notation