

EVALUATION CONTINUE ET EXAMENS.

Précis de docimologie.

Gilbert
Professeur

DU MEME AUTEUR

Introduction à la recherche en éducation, Paris, A. Colin-Bourrellier ; Liège, Thone, 1970, 3^e éd. revue et augmentée.

Einführung in die pädagogische Forschung, Weinheim, Julius Beltz Verlag, 1969.

La investigación pedagógica, Buenos Aires, Angel Estrada, 1971.

Les tests de connaissances, Bruxelles, Editest, 1965.

Rendement de l'enseignement des mathématiques dans douze pays (en collaboration avec T.N. POSTLETHWAITE), Paris, Institut Pédagogique National, 1969.

Comment les maîtres enseignent. Analyse des interactions verbales en classe, (avec la collaboration de E. BAYER), Bruxelles, Ministère de l'Education Nationale, Administration des Etudes, 1969.

H. BENJAMIN, *La pédagogie paléolithique ou Préhistoire de la Contestation*. Préface et adaptation française, Collection « Education 2000 », Paris, F. Nathan ; Bruxelles, Labor, 1970.

EVALUATI
ET I

PRECIS D

DEUXIEME EDI

Editions L A B O R
Rue Royale, 342
1030 BRUXELLES

Gilbert DE LANDSHEERE
Professeur à l'Université de Liège

EVALUATION CONTINUE ET EXAMENS

PRECIS DE DOCIMOLOGIE

DEUXIEME EDITION REVUE ET AUGMENTEE

Editions L A B O R
Rue Royale, 342
1030 BRUXELLES

Fernand NATHAN
Editeur
PARIS

INTRODUCTION.

De nouveaux modes d'évaluation des élèves entrent dans nos mœurs pédagogiques. Les résultats de l'*observation continue* prennent progressivement le pas sur l'examen de fin d'année.

L'innovation est louable. Elle ne traduit nullement la volonté de supprimer la *mesure* à l'école, mais, au contraire, de la rendre moralement plus juste et scientifiquement plus exacte.

Ce changement déborde largement le domaine de la notation et coïncide avec des transformations profondes de l'éducation. L'événement n'est pas fortuit. La civilisation contemporaine, l'économie de notre société réclament un homme pourvu de qualités et de connaissances nouvelles : en pareil cas, les innovations pédagogiques s'imposent irrésistiblement.

La contestation des examens traditionnels a toutefois créé un malentendu grave. Il ressemble de façon frappante à celui qui a surgi, entre 1920 et 1940 surtout, lorsque l'adoption des idées, mal digérées, de la pédagogie fonctionnelle des Claparède, Dewey, et autres « progressistes » conduisit au culte de l'improvisation, au mépris de la discipline rigoureuse, au centrage sur des intérêts n'existant que dans l'esprit des théoriciens...

Avec l'avènement de J. Dewey, l'école aurait pu prendre comme devise : « L'effort est mort, vive l'effort ! ». De même, serions-nous tenté de proposer aujourd'hui : « Les examens meurent, les examens sont morts, vivent les examens ! ».

Assurément, les vieux examens qui empoisonnaient l'atmosphère et la matière de cycles entiers d'études et décidaient de la carrière scolaire ou professionnelle en quelques heures, — voire en quelques minutes ! — doivent disparaître. Assurément aussi, la majeure partie des travaux docimologiques publiés jusqu'ici n'ont été que mises au point de systèmes

perfectionnés pour continuer à mal faire les choses. Car ils ne s'appuyaient pas sur une remise en cause de tout l'enseignement.

Mais si, par contre, on ménage, dans chaque activité d'enseignement de base, la place qui revient à l'évaluation, au *feedback* comme disent les psychologues d'aujourd'hui, les examens existent alors, de façon *quasi* permanente dans la vie scolaire. Tantôt ils consistent en brèves évaluations, tantôt ils prennent la forme d'épreuves plus longues, couvrant parfois des quantités considérables de matières.

L'essentiel est qu'ils ne s'insèrent pas, en corps étrangers, dans le processus d'éducation, mais qu'ils en fassent partie intégrante. Pour cette même raison d'ailleurs, ils ne se cantonnent donc pas étroitement au domaine de la connaissance, mais visent à saisir l'ensemble de la personnalité.

Toute réforme profonde s'accompagne, presque nécessairement, d'une période d'adaptation, de tâtonnements, d'errements. Ainsi s'explique le désarroi de beaucoup de maîtres qui essaient actuellement le système d'évaluation continue.

Ainsi s'explique aussi qu'un tabou frappe soudain le mot examen et, qu'au moment où nous écrivons ces lignes, on délivre çà et là des certificats de fin d'études sur la base d'évaluations ponctuelles, qualitativement et quantitativement insuffisantes, mal intégrées de surcroît, et sur la foi d'impressions personnelles à côté desquelles les vieux examens napoléoniens semblent des monuments d'objectivité ! Mais il ne s'agit, nous en sommes convaincu, que d'une crise d'adaptation.

Il n'est donc nullement paradoxal qu'après avoir recommandé — brièvement — le procès des examens et nous être déclaré ardent partisan de l'observation continue et du bilan de fin de cycle, nous consacrons le présent ouvrage... à l'organisation des examens. Mais le lecteur sait maintenant que, si le mot est resté le même, son sens a profondément changé.

*

**

Beaucoup d'enseignants se méfient encore des tests et des autres instruments qui ambitionnent de quantifier l'humain. Non sans raison d'ailleurs, car ces outils ont fait, dans les dernières décennies, leur maladie de jeunesse. Et comme leurs utilisateurs n'ont pas toujours su compenser l'imperfection des techniques par une grande modération de jugement et une intelligence profonde des situations d'ensemble, bien des erreurs ont été commises.

Mais si la prudence et le sens clinique ne sont pas assez considérables ont maintenant mesurer bien satisfaisante.

Aujourd'hui, d'aucuns ne plus parce que tout ce qu'ils n'en comprennent ce qu'on ignore, est un mathématique des méthodes expérimentale, souvent indigestion.

Le temps de l'opposés de l'éducation est leurs chiffres ; les seconds leurs procédures ; tous efforts pour le plus grand.

Que ceux qui, en quelques chiffres se rassureront un bagage suffisant.

Comme d'habitude, sont plus difficiles à expliquer.

Certes, tous les seconds ne seront pas livrés, mais rencontrer ont une valeur en outre, un indéniablement ultérieurement la lecture spécialisée.

ire les choses. Car ils ne
cause de tout l'enseigne-

ans chaque activité d'en-
vient à l'évaluation, au
es d'aujourd'hui, les exa-
permanente dans la vie
es évaluations, tantôt ils
ues, couvrant parfois des

pas, en corps étrangers,
qu'ils en fassent partie
ailleurs, ils ne se canton-
ine de la connaissance,
a personnalité.

agne, presque nécessai-
de tâtonnements, d'erre-
beaucoup de maîtres qui
évaluation continue.

u frappe soudain le mot
ivons ces lignes, on déli-
des sur la base d'évalua-
quantitativement insuffisan-
foi d'impressions person-
mens napoléoniens sem-
ais il ne s'agit, nous en
d'adaptation.

l qu'après avoir recom-
s examens et nous être
n continue et du bilan de
sent ouvrage... à l'orga-
sait maintenant que, si
a profondément changé.

ent encore des tests et
nt de quantifier l'humain.
outils ont fait, dans les
eunesse. Et comme leurs
enser l'imperfection des
de jugement et une intel-
semble, bien des erreurs

Mais si la prudence reste nécessaire, si l'esprit critique
et le sens clinique ne perdront jamais leurs droits, des pro-
grès considérables ont néanmoins été accomplis : on peut
maintenant mesurer bien des comportements humains de façon
satisfaisante.

Aujourd'hui, d'aucuns pourraient rejeter la mesure objective,
non plus parce que toute validité lui fait défaut, mais parce
qu'ils n'en comprennent pas l'économie. Dénier toute valeur à
ce qu'on ignore, est un réflexe de défense bien connu. L'aspect
mathématique des méthodes à adopter et de la littérature expé-
rimentale, souvent indigeste, vient encore compliquer la situa-
tion.

Le temps de l'opposition entre psychométriciens et prati-
ciens de l'éducation est révolu. Les premiers doivent humaniser
leurs chiffres ; les seconds, introduire plus de rigueur dans
leurs procédures ; tous deux doivent unir et harmoniser leurs
efforts pour le plus grand profit de l'étudiant et de la commu-
nauté.

*
**

Que ceux qui, en feuilletant ce livre, ont aperçu quel-
ques chiffres se rassurent : les quatre opérations arithmétiques
seront un bagage suffisant !

Comme d'habitude, les procédés dont il va être question
sont plus difficiles à expliquer qu'à appliquer.

Certes, tous les secrets de la mesure et de la statistique
ne seront pas livrés, mais les notions simples que nous allons
rencontrer ont une valeur pratique éprouvée. Elles détiennent,
en outre, un indéniable pouvoir de démythification et faciliteront
ultérieurement la lecture de travaux docimologiques plus spé-
cialisés.

PREMIERE PARTIE

DEFINITIONS

*Les définitions sont données
à la page 10.*

I. DOCIMOLOGIE, DOCIMASTIQUE ET DOXOLOGIE.

La *docimologie* est une science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés.

La *docimastique* est la technique des examens.

Au début, la docimologie a revêtu un caractère négatif en critiquant les modes de notation et en montrant expérimentalement le manque de fidélité et de validité des examens.

Par la suite, elle est entrée dans une phase constructive en essayant de proposer des méthodes et des techniques de mesure plus objectives ou, au moins, plus rigoureuses, et en mettant au point les moyens de rendre les notes comparables, de façon à assurer plus de justice scolaire.

La *doxologie* est l'étude systématique du rôle que l'évaluation joue dans l'éducation scolaire.

J. Guillaumin¹ lui assigne, notamment, les objectifs suivants :

- Etude des effets inhibiteurs ou stimulants des différentes formes d'examens ;
- Etude des réactions émotionnelles des élèves et, de là, des réactions intellectuelles aux jugements du maître ;
- Etude des effets de l'opinion du maître concernant les élèves sur son enseignement et sur l'apprentissage scolaire ;
- Etude des processus mis en jeu et des effets obtenus par l'automatisation, par l'internotation, par la notation d'équipe, par l'absence de notation.

¹ J. GUILLAUMIN, L'aspect interpersonnel de la notation : de la docimologie à la doxologie pédagogique, in *Bulletin de la Société A. Binet et T. Simon*, 86, 1968, 250-275.

COURS. ON CONTINUES.

illement dissociables de
r synonyme après avoir
e, le malheur, le danger

déterminée par une note
passer, tandis que dans
osées est fixé d'avance.
celui du refus, est, dans
la procédure soit enta-
mmes pas très éloignés

examen, nous percevons
e des concepts d'obser-
rénité, la bienveillance,
être qui suit, avec une
s exclue, le long chemi-
du moment et la pleine

d'étapes ; les concours
es points dans la durée,
l'éducation.

événements consiste fon-
nuelle de comportements
tions éclairant sur leur
ppartient pas d'engager
s et des effets du renfor-
t que, sans lui, l'appren-
duire.

nous lui donnons dans
place importante dans
égrante. Elle a toujours,
vec le progrès, en exten-

ourvu des qualités intel-
connaissances nécessai-
lle ou un cycle d'études
e trouver ? Répondre à
succès dans l'étape qui

2° Un rôle de jaugeage :

- a) Contrôle des acquisitions ;
- b) Evaluation du progrès, cas où l'on compare l'élève à lui-même ;
- c) Situation de l'élève à un moment donné :
 - dans sa classe ou son groupe de travail ?
 - dans l'ensemble des classes parallèles d'une même école ?
 - dans des ensembles plus vastes : ville, canton, province, pays ?

Il ne s'agit pas nécessairement de procéder à un examen ou à un concours, mais de faire le point, de déterminer la position relative.

3° Un rôle diagnostique :

Pourquoi un apprentissage parfait ne s'est-il pas produit ? Quelles matières ou techniques l'étudiant domine-t-il insuffisamment, quels sont les processus mentaux en cause ?

Les instruments nécessaires à l'évaluation, à l'examen ou au concours ne diffèrent pas toujours, mais bien la façon de les utiliser. Aussi, pour éviter des précisions verbales fastidieuses emploierons-nous, dans les pages qui suivent, le mot examen dans deux sens différents, que le contexte éclairera toujours : examen proprement dit et, plus généralement, toute procédure pédagogique ayant une mesure d'apprentissages ou de connaissances pour objet.

Examens internes et examens externes.

Au sens strict, l'*examen interne* dans une branche est construit et noté par le maître qui l'a enseignée, et subi par les élèves qui ont reçu cet enseignement, dans le cadre de la classe ou de l'école.

Au sens plus large, on qualifie d'internes les examens organisés indépendamment dans chaque école, qu'il existe ou non une coordination ou une unification par branche et par niveaux et sections.

Par *examens externes*, on désigne les épreuves organisées et notées par des jurys indépendants des écoles, à l'échelon local, régional ou national. Les plus connues de ces épreuves sont celles du baccalauréat français ; citons encore, en Belgique, les examens cantonaux, en fin d'études primaires, et, en Angleterre, jusqu'à ces derniers temps, l'*Eleven + Examination* à l'entrée dans l'enseignement secondaire.

III. NOTES ET SCORES.

La distinction entre notes et scores rendrait, pensons-nous, un grand service en éducation.

Dans une dictée, l'élève peut commettre un certain nombre de fautes dont il ne nous appartient guère d'évaluer l'existence : elles sont ou ne sont pas. Toutefois, le nombre de fautes relevées n'a, en soi, aucune signification éducative : faire cinq fautes dans la dictée de Mérimée témoigne d'une extraordinaire connaissance des arcanes de l'orthographe ; en d'autres conditions, le même résultat annonce une faiblesse grave. Une information relativiste est donc aussi nécessaire.

Par *score*, nous désignerons les résultats objectifs obtenus à un test ou à toute autre forme d'évaluation par *compte ou décompte de points, selon des règles fixes* : nombre de fautes en dictées, résultats à un test standardisé.

Par *note*, nous entendrons une appréciation synthétique traduisant l'évaluation d'une performance dans le domaine de l'éducation.

La note peut être objective ou subjective, mais elle est toujours relative. Attribuer la note A à un élève dont la performance se situe à tel niveau dans un étalonnage national relève de la première catégorie ; marquer sa composition d'un *bien* relève de la seconde.

DEUXIÈME

L'ACC
ET LA

CORES.

res rendrait, pensons-nous,

mettre un certain nombre
uère d'évaluer l'existence :
le nombre de fautes rele-
ion éducative : faire cinq
noigne d'une extraordinaire
graphe ; en d'autres condi-
faiblesse grave. Une infor-
ssaire.

résultats objectifs obtenus
évaluation par *compte ou*
s fixes : nombre de fautes
disé.

ppréciation synthétique tra-
e dans le domaine de l'édu-

bjective, mais elle est tou-
élève dont la performance
nage national relève de la
mposition d'un *bien* relève

DEUXIEME PARTIE

L'ACCUSATION ET LA DEFENSE

Une critique détaillée par H. Piéron dans son ouvrage¹. Tout enseignant doit

Le présent chapitre comporte quelques-unes sont de brefs rappels sommairement mises en lumière concernant des phénomènes tels que l'effet de halo, effet oedipien, l'examen est en cause ; tout cela dans la quelle il s'insère.

1. Corps étrangers dans le service d'une pédagogie

Qu'il s'agisse d'intermittents trimestriels ou annuels, le contrôle de rétention de connaissances ne concerne pas seulement les aspects les plus évidents, aussi à peu près tous les aspects de la rétention bien comprise doit cultiver

Nous nous trouvons, dans ce système pédagogique où les leçons servent de banc d'essai et de valeurs, sélectionnées à jouer dans une société

Dans ce cadre, l'examen est de fabrication, de vérification, de gabarit, bref aux spécificités

L'épanouissement de l'élève trouve de place dans ces préoccupations conçue pour elle, mais pour

¹ Paris, P.U.F., 1963.

CHAPITRE 1.

CRITIQUE DES EXAMENS.

Une critique détaillée des examens traditionnels a été faite par H. Piéron dans son excellent ouvrage *Examens et docimologie*¹. Tout enseignant doit l'avoir lu.

Le présent chapitre contient deux types de remarques. Les unes sont de brefs rappels d'imperfections bien connues, inlassablement mises en lumière par les docimologistes ; les autres concernent des phénomènes moins bien étudiés : stéréotypie, effet de halo, effet oedipien de la prédiction, ... Tantôt, seul l'examen est en cause ; tantôt, c'est toute la pédagogie dans laquelle il s'insère.

1. Corps étrangers dans l'éducation, au service d'une pédagogie dépassée.

Qu'il s'agisse d'interrogations périodiques ou d'examens trimestriels ou annuels, l'évaluation se réduit souvent à un contrôle de rétention de connaissances laissant inexplorés, non seulement les aspects les plus importants de l'intelligence, mais aussi à peu près tous les traits de personnalité qu'une éducation bien comprise doit cultiver.

Nous nous trouvons, en fait, devant les séquelles d'un système pédagogique où, comme T. Brameld aime à le dire, les leçons servent de bandes transporteuses de connaissances et de valeurs, sélectionnées en fonction d'un rôle prédéterminé à jouer dans une société non démocratique.

Dans ce cadre, l'examen constitue une sorte de contrôle de fabrication, de vérification de la conformité au moule, au gabarit, bref aux spécifications imposées par l'autorité.

L'épanouissement de la personne humaine n'occupe guère de place dans ces préoccupations, car l'éducation n'est pas conçue pour elle, mais pour le service d'un régime.

¹ Paris, P.U.F., 1963.

es de cette situation ont été
servir l'élève, de l'informer
es comportements en cours
une adaptation meilleure,
ptée au point que l'étudiant
ément, l'examen est aussi
l'impôt qui, elle non plus,
de civisme.

a procédure combinant son
des questions posées con-
table : dans les quinze mois
0 % des matières factuelles

sociale jouée par l'examen,
qué, ni instruit.

ssite scolaire conditionne la
point que M. Young a cru
a « méritocratie », l'examen
u de l'obtention du diplôme
e. Même en cours d'année,
ssances et d'aptitudes sont
ion, ce qui n'est certes pas
fondément le rôle éducatif

dition veut que, chez nous,
ur toutes les branches se
ques jours. Lors d'une ses-
e, le triste privilège d'être
eux jours de suite... Dans
ne phénomène de concen-
cours d'année, lorsque l'ap-
bulletin périodique suscite

ation des épreuves et leur
r l'apport de matières nou-
s professeurs ne respectent
on ou ne les exploitent pas
ignent considérablement le

conditions difficiles revêtent
: la vie moderne ne nous
d'y être préparé. D'accord,
sent dans la clarté !

Avant d'organiser un examen, il importe de définir claire-
ment son but : s'agit-il de contrôler l'acquisition des connais-
sances ? la résistance au stress ? ou la capacité de restituer et
d'user des connaissances en situation de stress ? Ces trois
objectifs diffèrent et appellent des épreuves différentes.

3. Inégalité - injustice.

Dans notre système scolaire, les professeurs rédigent cha-
cun les questions d'examen destinées à leurs élèves. Le prin-
cipe est excellent et dans des pays comme la Grande-Bretagne
où, traditionnellement, tous les examens-clés étaient confiés à
des commissions extérieures à l'école, on tend à accorder à
chaque enseignant la responsabilité du contrôle. Toutefois,
les méthodes anglo-saxonnes présentent plus de rigueur que
les nôtres, car des mesures précises sont prises pour assurer
la comparabilité des examens dans la région, voire dans tout
le pays. Nous consacrons d'ailleurs une partie de ce livre aux
procédés utilisés pour atteindre ce but.

La liberté, à peu près totale, laissée à nos maîtres conduit
à des situations injustes.

Les écoles de notre pays constituent des mondes isolés
dont les populations présentent des caractéristiques parfois
fort différentes. Tel petit établissement d'enseignement secon-
daire ne compte que quelques élèves, en majorité handicapés
par leurs origines socio-économiques. A l'opposé, tel autre
établissement situé dans une grande ville voit ses sections
« fortes » peuplées d'une majorité favorisée socialement et intel-
lectuellement, les adolescents en difficulté étant orientés vers
une école voisine, qui a la réputation d'être plus compréhensive.

Avec le temps, et la concurrence aidant, des traditions de
sévérité ou de « générosité » se créent dans les établissements.
Marion Coulon observait récemment que, quand des écoles de
même type sont successivement créées dans une même région,
la sévérité des examens est inversement proportionnelle à l'an-
cienneté de l'établissement. Nous ne pouvons certes pas géné-
raliser, mais qui oserait affirmer que la proposition est dénuée
de tout fondement ? Beaucoup de parents l'ont compris, qui
ambitionnent pour leur enfant un diplôme déterminé, sans trop
se soucier de la valeur intrinsèque des études : en cas d'échec,
il est plus aisé de changer d'école que d'installer un traitement
ou de modifier l'orientation.

Quoi qu'il en soit, les maîtres adaptent leur enseignement
et les examens au niveau de leur classe (ce qui est louable),
avec la conséquence qu'un élève bien noté dans une popula-
tion faible serait parmi les très faibles ou échouerait dans un
groupe fort.

L'effet cumulatif de tous les ajustements circonstanciels aboutit parfois à des situations scandaleuses. Il y a quelques années, nous avons pu consulter des statistiques établissant que, pour un même type et un même niveau d'enseignement secondaire, les chances de réussite d'un élève variaient de un à huit selon l'école fréquentée. Le cas est probablement extrême, encore manquons-nous d'études précises permettant de savoir ce qui se passe réellement.

Il faut y insister, l'adaptation du maître au niveau de ses élèves n'est pas, en soi, critiquable ; elle est, au contraire, caractéristique du bon éducateur.

L'injustice apparaît lorsque des notes purement relatives sont utilisées en situation de concurrence extérieure, ou sont l'unique critère pour l'obtention d'un diplôme dont peut dépendre l'avenir des élèves.

De toute façon, on constate des variations considérables dans la *quantité des matières couvertes* et dans la *qualité des réponses exigées*. Ici, la restitution de mémoire d'une partie d'un court syllabus suffit, alors que là, les questions exigent analyse, synthèse et jugement personnel à propos d'une matière abondante.

Tel maître ne pose qu'une ou deux questions, plus choisies pour la facilité de correction qu'elles offrent que pour leur importance réelle, tel autre s'efforce de parcourir tout le cours, au risque de transformer l'examen en un marathon. En sport, on sait que cette longue course doit être réservée à des individus exceptionnels ; pourquoi, dans le domaine de l'instruction, les trente élèves d'une classe seraient-ils tous des coureurs de fond ?

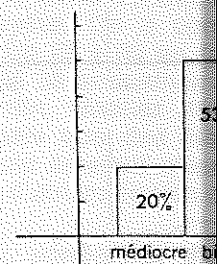
Avant de poursuivre, trois résultats de recherches nous permettront d'illustrer les phénomènes dont il vient d'être question.

R. Gjorgjevski¹ a montré, par une expérience simple, comment, par relativité, des professeurs évaluent différemment une même performance.

Cinq correcteurs, professeurs de la même branche, ont noté indépendamment cent épreuves écrites provenant d'une école secondaire. On a ensuite extrait quinze copies qui toutes avaient obtenu la note « bien ». Elles ont été confiées, pour

¹ Voir N. ROT et Z. BUTAS, Les distributions des notes scolaires comparées aux distributions des résultats obtenus aux tests de connaissances, in *Le travail humain*, XXII, 1-2, 1959.

nouvelle correction, à quatre-vingt-cinq pour cent, spontanément adopté des ex la moitié des cas environ, la répartition nouvelle¹ :



Poursuivant son expérience, le même groupe de cent copies a été corrigé par un groupe de bons, et un groupe de do

Dans chaque groupe, il y avait cinquante copies. Deux fois cinq professeurs ont corrigé les copies glissées dans les groupes de 3 sur 5 à 2,40 ; dans les groupes de 3 à 3,87.

On doit, par ailleurs, à propos de 4 860 élèves répartis dans 30 groupes, 30 instituteurs ont été invités à classer les copies en cinq catégories : très bons, bons, moyens, faibles, et très faibles. On a fait passer un même test à tous les élèves. Le dépouillement des copies a été fait à un degré de connaissances était jugé très bons, bo

Comme le souligne peut-être, on peut dire (...) que l'instituteur, en notant les points du programme, classe chacun de ses élèves. Ce qui est en fait une faute à chaque faiblesse, à moins qu'il ne s'agisse d'un élève non plus au sein d'un groupe. On a noté des divergences d'appréciation

¹ Le nombre de copies est trop petit pour une signification statistique.

² Voir *Le Travail humain*, XXII, 1

ajustements circonstanciels
andaleuses. Il y a quelques
des statistiques établissant
ême niveau d'enseignement
d'un élève variaient de un à
s est probablement extrême,
écises permettant de savoir

du maître au niveau de ses
ble ; elle est, au contraire,

es notes purement relatives
urrence extérieure, ou sont
n diplôme dont peut dépen-

es variations considérables
ertes et dans la *qualité des*
n de mémoire d'une partie
e là, les questions exigent
sonnel à propos d'une ma-

eux questions, plus choisies
elles offrent que pour leur
e de parcourir tout le cours,
en un marathon. En sport,
it être réservée à des indi-
le domaine de l'instruction,
aient-ils tous des coureurs

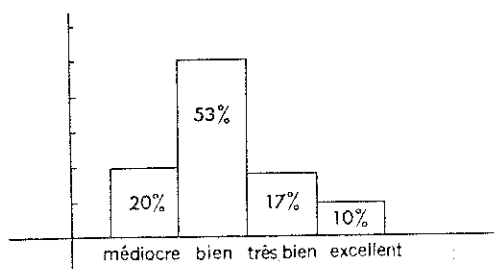
ultats de recherches nous
ènes dont il vient d'être

r une expérience simple,
eurs évaluent différemment

de la même branche, ont
es écrites provenant d'une
ait quinze copies qui toutes
es ont été confiées, pour

s des notes scolaires comparées aux
connaissances, in *Le travail humain*,

nouvelle correction, à quatre autres professeurs. Ceux-ci ont
spontanément adopté des exigences nouvelles : seulement dans
la moitié des cas environ, la note est restée « bien ». Voici la
répartition nouvelle ¹ :



Poursuivant son expérience, Gjorgjevski a extrait, de la
même série de cent copies, un groupe de douze jugées très
bonnes, et un groupe de douze jugées médiocres.

Dans chaque groupe, il a glissé trois copies jugées bonnes.
Deux fois cinq professeurs ont évalué. La moyenne des « bon-
nes » copies glissées dans les « très bonnes » est descendue
de 3 sur 5 à 2,40 ; dans le groupe opposé, la moyenne est
passée de 3 à 3,87.

On doit, par ailleurs, à M. Reuchlin ² une étude portant sur
4 860 élèves répartis dans 397 écoles primaires de France. Les
institutrices ont été invitées à classer leurs élèves en quatre
catégories : très bons, bons, moyens, médiocres. On a ensuite
fait passer un même test de français et de calcul à tous les
élèves. Le dépouillement des résultats a montré que le même
degré de connaissances était le fait d'élèves qui, selon l'école,
étaient jugés très bons, bons, moyens ou médiocres.

Comme le souligne pertinemment M. Reuchlin : « On peut
dire (...) que l'institutrice, certainement, connaît mieux que per-
sonne les points du programme qui sont acquis ou non par
chacun de ses élèves. Ce qu'il ignore, c'est la gravité qui s'atta-
che à chaque faiblesse, à chaque lacune, lorsqu'on la consi-
dère non plus au sein d'une classe qui peut être « forte » ou
« faible », mais par rapport à l'ensemble du pays. De là, les
divergences d'appréciation mises en lumière par l'enquête. »

¹ Le nombre de copies est trop peu élevé pour conduire à des résultats ayant une
signification statistique.

² Voir *Le Travail humain*, XXII, 1-2, 1959, p. 12 sq.

Les graphiques ci-dessous synthétisent les résultats.

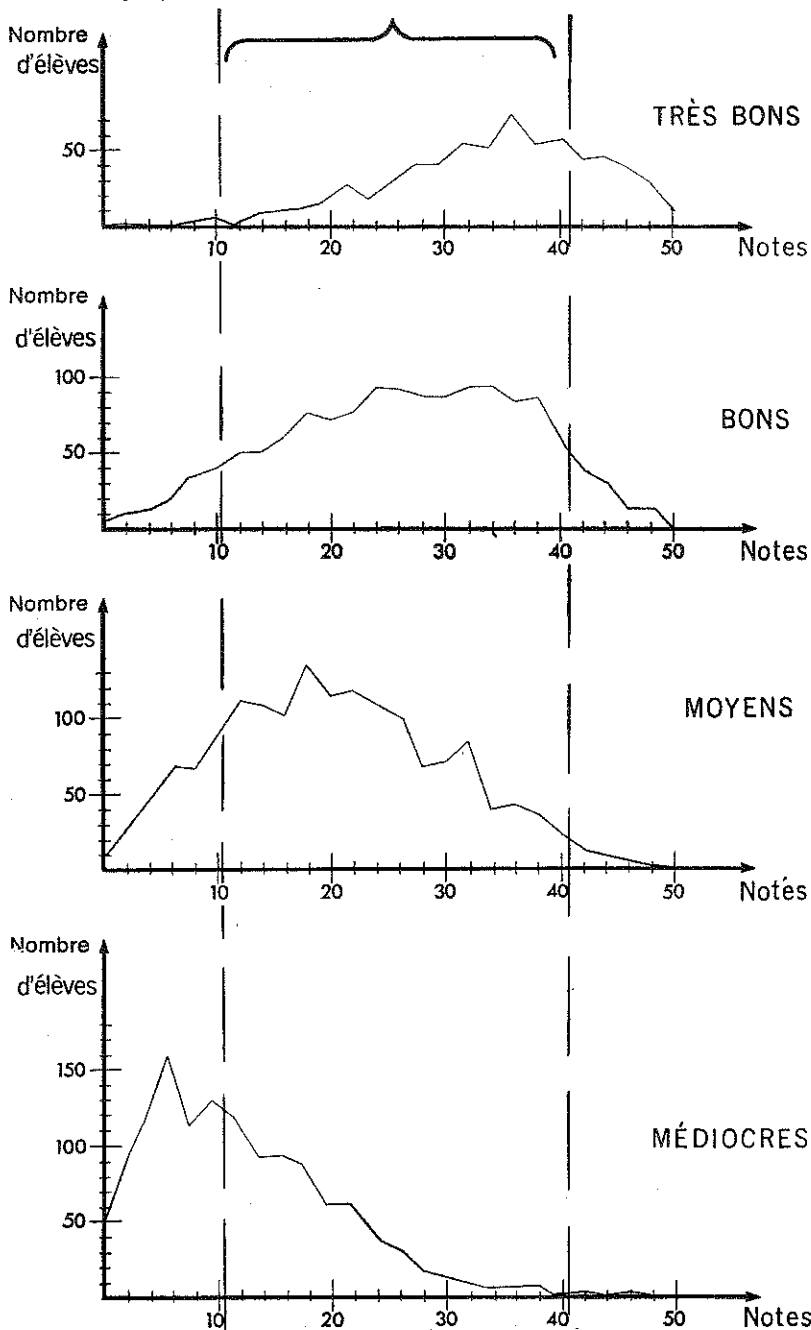


Fig. 1. — Les quatre graphiques permettent de comparer la distribution des notes obtenues, dans l'épreuve de calcul, par 654 élèves jugés « très bons » par leurs maîtres, par 1 303 élèves « bons », 1 551 « moyens » et 1 300 élèves jugés « médiocres ». On voit que les distributions de notes se recouvrent largement : dans la zone de notes qui va de 10 à 40, le même niveau de connaissances peut être qualifié, selon le cas, de très bon, moyen ou médiocre.

Voici encore deux ex

« Dans une enquête fra représentatif d'élèves de troi daire), on a fait passer une res et une épreuve de conna de 406 classes. On a cons s'étaient de 23 à 60 en fra 80) et de 7 à 38 en mathéma 44) (Reuchlin, Bacher 196 (1964) signale que des élév de leur école seraient classa quentaient d'autres écoles de

4. L'échec, générateur d'é

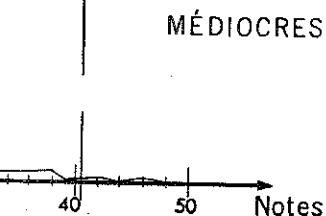
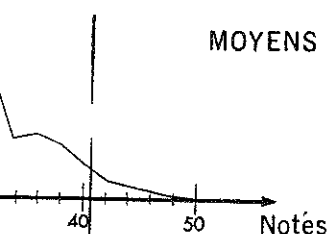
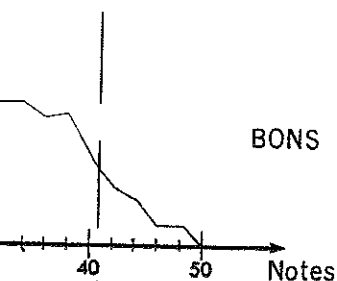
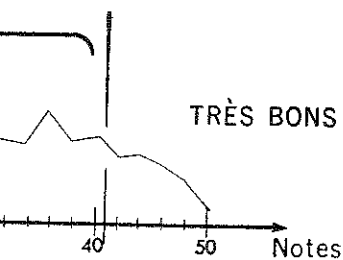
On peut objecter que l chargeront de rétablir la ju raient pas respectée. Les ch aussi simples. D'abord parc ragés injustement de l'ens exemple, n'auront peut-être chance à l'université. Ensu de gagner sur tous les table

Les recherches actuelle gique : le succès engendre prépare un échec demain. P valorisent un capital intelle « à bon marché » a parfoi de ne pas pousser les par où il se trouve et de perm sorte de nouveau souffle. Ce daire ne voient-ils pas le re ter spectaculairement dès gique de l'adolescence pe

R. Rosenthal et L. Jac étude d'ensemble sur l'asp tion, ainsi qualifié parce qu n'avait pas prédit qu'il tu éloigné de sa famille. Con tué. En d'autres mots, la avait été prédite.

1 F. BACHER, La normalisation d spécial de la revue *Les Sciences d*
2 *Pvgmalion in the Classroom, Te ment*, New York, Holt, Rinehart and W

mathématisent les résultats.



comparer la distribution des notes jugés « très bons » par leurs maîtres, élèves jugés « médiocres ». On voit nettement : dans la zone de notes qui va de 40 à 50, la courbe est qualifiée, selon le cas, de très bon,

Voici encore deux exemples donnés par F. Bacher¹ :

« Dans une enquête française portant sur un échantillon représentatif d'élèves de troisième (fin du premier cycle secondaire), on a fait passer une épreuve de connaissances littéraires et une épreuve de connaissances mathématiques aux élèves de 406 classes. On a constaté que les moyennes de classe s'étaient de 23 à 60 en français (pour une épreuve notée sur 80) et de 7 à 38 en mathématiques (pour une épreuve notée sur 44) (Reuchlin, Bacher 1968). (...) Aux Etats-Unis, Flanagan (1964) signale que des élèves classés dans le quart inférieur de leur école seraient classés dans le quart supérieur s'ils fréquentaient d'autres écoles de la même région. »

4. L'échec, générateur d'échecs.

On peut objecter que les études supérieures ou la vie se chargeront de rétablir la justice que les notes scolaires n'auraient pas respectée. Les choses ne sont malheureusement pas aussi simples. D'abord parce que les élèves écartés ou découragés injustement de l'enseignement secondaire général, par exemple, n'auront peut-être plus l'occasion de défendre leur chance à l'université. Ensuite parce que les favoris risquent de gagner sur tous les tableaux.

Les recherches actuelles confirment un vieil adage pédagogique : le succès engendre le succès et l'échec d'aujourd'hui prépare un échec demain. Portés par la réussite, certains élèves valorisent un capital intellectuel limité. Le passage de classe « à bon marché » a parfois la double conséquence salutaire de ne pas pousser les parents à retirer leur enfant de l'école où il se trouve et de permettre à l'étudiant de retrouver une sorte de nouveau souffle. Combien de fois les maîtres du secondaire ne voient-ils pas le rendement scolaire d'un élève remonter spectaculairement dès que la crise physique et psychologique de l'adolescence perd de son acuité ?

R. Rosenthal et L. Jacobson² ont récemment publié une étude d'ensemble sur l'aspect de l'*effet œdipien de la prédiction*, ainsi qualifié parce que si, à la naissance d'Œdipe, l'oracle n'avait pas prédit qu'il tuerait son père, il n'aurait pas été éloigné de sa famille. Connaissant son père, il ne l'aurait pas tué. En d'autres mots, la tragédie se produisit *parce qu'elle avait été prédite*.

¹ F. BACHER, La normalisation de la notation, in *Docimologie et Education*, numéro spécial de la revue *Les Sciences de l'Education*, 2-3, 1969, pp. 51-52.

² *Pygmalion in the Classroom, Teacher Expectations and Pupil's Intellectual Development*, New York, Holt, Rinehart and Winston, 1968.

Il semble que, dans une mesure certes difficile à apprécier, un élève se comporte en fonction du jugement que le professeur porte sur lui. Même les animaux en dressage n'échappent pas à cette règle. Rosenthal rappelle l'expérience suivante : on constitue deux groupes de souris blanches génétiquement identiques. Au moment où on les remet aux étudiants chargés de les dresser, une remarque indique que le premier groupe est composé d'animaux particulièrement bien doués, alors que le second est de pauvre qualité. Les résultats du dressage confirmeront ce pronostic... fantaisiste.

L'expérience suivante, conduite par Rosenthal lui-même, s'inscrit dans le même contexte.

On a annoncé à des instituteurs des six années primaires de l'Oak School que d'éminents chercheurs venaient de mettre au point un test d'« épanouissement » (en réalité, un simple test d'intelligence, ¹ peu connu et sans vertu particulière, a été utilisé).

Tous les élèves de l'école y ont été soumis, et on a signalé ceux qui étaient sur le point de s'épanouir intellectuellement, pronostic fantaisiste.

Le quotient intellectuel des élèves ainsi désignés s'est élevé d'une façon significative dans les trois années inférieures ; en outre, des progrès supérieurs à la moyenne ont été observés en lecture et en arithmétique ².

5. Rupture entre enseignement et examen.

L'examen doit être le reflet de l'enseignement donné. Imaginons qu'au lieu de conduire ses élèves à la piscine, un professeur d'éducation physique ait passé son année à disserter sur la natation. Il serait évidemment inadmissible que les notes d'examen soient attribuées d'après les performances natatoires... car seuls réussiraient ceux qui ont appris à nager en dehors du cours.

Pareil agissement paraît impensable. Pourtant, combien de fois n'assiste-t-on pas à une aberration similaire. Le professeur fait un cours où il impose les informations, les opinions et les jugements, puis, à l'examen, il pose des questions dites d'intel-

¹ Mesuré à l'aide du Toga de Flanagan (1960).

² Les chiffres avancés par Rosenthal et Jacobsen sont contestés par R.L. Thorndike. Par contre, la tendance à l'augmentation des QI. et des scores semble reconnue par tous les spécialistes. C'est évidemment ce qui importe. Pour plus de détails, voir *American Educational Research Journal*, 5, 4, 1968, 708.

ligence, entendons qui me n'ont précisément pas été : quoi les étudiants seraient seuls, à l'examen, la solutionales, s'ils n'ont pas pu le

Ainsi s'expliquent bien maux. Une enquête suscit insuffisantes chez un prof avaient obtenu plus de la (d'année) fit rapidement app que deux ou trois interrog questions d'examen, deux fert, démarches qui, au m plaçait, nécessitaient une c mènes et un entraînement, tion de problèmes originaux

Une discussion amicale révéla qu'il se croyait obli points du programme, qu'il site de l'entraînement systé lisé à la solution de problèr il ne possédait pas les no nécessaires à la conduite d

6. Désaccord entre corre

Tous les docimologist divergences entre maîtres d'élève ou un même ense tant une relativité plus sù même professeur jugeant u de temps (fidélité) peuv

Comme tous les ou d'exemples, nous n'en p rappel.

a) *La composition françai*

Elle est une accusée d

- Piéron rapporte qu'une r recteurs, tous professe aux résultats suivants ¹ :

¹ PIERON, *o.c.*, p. 123.

ertes difficile à apprécier,
u jugement que le profes-
en dressage n'échappent
le l'expérience suivante :
s blanches génétiquement
net aux étudiants chargés
e que le premier groupe
ement bien doués, alors
Les résultats du dressage

par Rosenthal lui-même,

des six années primaires
cheurs venaient de mettre
t» (en réalité, un simple
ns vertu particulière, a été

été soumis, et on a signalé
panouir intellectuellement,

èves ainsi désignés s'est
s les trois années infé-
érieurs à la moyenne ont
étique².

xamen.

enseignement donné. Im-
èves à la piscine, un pro-
ssé son année à disserter
inadmissible que les notes
les performances natatoi-
ui ont appris à nager en

able. Pourtant, combien de
on similaire. Le professeur
ations, les opinions et les
des questions dites d'intel-

en sont contestés par R.L. Thorndike
et des scores semble reconnue par
e. Pour plus de détails, voir *American*

ligence, entendons qui mettent en œuvre des capacités qui
n'ont précisément pas été installées pendant les leçons. Pour-
quoi les étudiants seraient-ils soudain capables de découvrir
seuls, à l'examen, la solution d'applications géométriques origi-
nales, s'ils n'ont pas pu le faire pendant l'année ?

Ainsi s'expliquent bien des pourcentages d'échecs anor-
maux. Une enquête suscitée à cause de nombreuses notes
insuffisantes chez un professeur de chimie (4 élèves sur 22
avaient obtenu plus de la moitié des points à l'examen de fin
d'année) fit rapidement apparaître que le professeur ne faisait
que deux ou trois interrogations écrites par an. Sur les trois
questions d'examen, deux impliquaient interprétation et trans-
fert, démarches qui, au niveau de difficulté où le maître se
plaçait, nécessitaient une compréhension profonde des phéno-
mènes et un entraînement, long et finement contrôlé, à la solu-
tion de problèmes originaux.

Une discussion amicale avec le jeune professeur en question
révéla qu'il se croyait obligé de traiter en détail de tous les
points du programme, qu'il n'avait pas conscience de la néces-
sité de l'entraînement systématique et au moins semi-indivua-
lisé à la solution de problèmes originaux et que, de toute façon,
il ne possédait pas les notions psychopédagogiques de base
nécessaires à la conduite d'un tel entraînement.

6. Désaccord entre correcteurs.

Tous les docimologistes citent des exemples de grandes
divergences entre maîtres appelés à évaluer un même travail
d'élève ou un même ensemble de travaux, ce qui permet pour-
tant une relativité plus sûre. Les fluctuations des notes d'un
même professeur jugeant un même travail à quelque intervalle
de temps (fidélité) peuvent aussi être importantes.

Comme tous les ouvrages de docimologie fourmillent
d'exemples, nous n'en proposons que quelques-uns, pour
rappel.

a) *La composition française.*

Elle est une accusée de prédilection.

- Piéron rapporte qu'une même composition, jugée par 76 cor-
recteurs, tous professeurs de langue maternelle, conduisit
aux résultats suivants¹ :

¹ PIERON, o.c., p. 123.

Note sur 20	Nombre de correcteurs qui l'ont attribuée
de 0 à 1	1
2 - 3	6
4 - 5	20
6 - 7	34
8 - 9	10
10 - 11	3
12 - 13	2

- Le Service de la Recherche pédagogique du C.R.D.P. de Lyon a invité, en 1967¹, 150 professeurs de français de l'Académie de Lyon à corriger trois compositions françaises traitant d'un même sujet.

Max. 20	Moyenne	Marge de variation
Devoir I	10 1/4	4 1/2 - 13 3/4
Devoir II	6,5	2 1/2 - 12 1/2
Devoir III	11,6	5 1/2 - 17 1/2

- L'intervention d'hommes cultivés, étrangers aux routines de l'enseignement améliorerait-elle la situation ? P. Diederich et ses collaborateurs² ont comparé la façon de corriger de 10 professeurs de langue maternelle et de 43 autres notateurs : spécialistes des sciences sociales, des sciences naturelles, écrivains, éditeurs, avocats et hommes d'affaires (cadres supérieurs).

Tous ont été invités, sans autre précision, à classer 300 compositions en 9 tas, par ordre de mérite général. Aucun des

1 C.R.D.P., *Docimologie et Examens*, Lyon, I.P.N., 1969.
 2 P. DIEDERICH et al., *Factors in Judgments of Writing Ability*, Princeton, E.T.S., Res. Bulletin RB-61-15, 1961.

tas ne pouvait compter n
 l'ensemble se sont vu att
 possibles. Aucune compo
 sements différents...

b) *Mathématiques.*

Le C.R.D.P. de Lyon l
 de mathématiques exerçan
 les copies de 3 élèves aya

Max. 20
Elève I
Elève II
Elève III

Si, pour l'élève I, par
 extrêmes, il reste autant d
 8/20, alors qu'on pouvait a

c) *Médecine.*

Au niveau universitaire

Un même candidat de
 septembre 1965, une épreu
 simples et précises à traite

Les réponses (anonym
 ment par cinq correcteurs
 identiques.

Résultat total sur 100	Ma
Copie de juin	
Copie de septembre	

1 O.c.
 Voir *Le Monde* du 6 septembre 1
 3 Remarquons que les trois adm

Nombre de correcteurs qui l'ont attribuée	
1	1
6	6
20	20
34	34
10	10
3	3
2	2

gique du C.R.D.P. de Lyon
de français de l'Académie
ns françaises traitant d'un

Marge de variation
$4\frac{1}{2} - 13\frac{3}{4}$
$2\frac{1}{2} - 12\frac{1}{2}$
$5\frac{1}{2} - 17\frac{1}{2}$

étrangers aux routines de
situation ? P. Diederich et
a façon de corriger de 10
t de 43 autres notateurs :
des sciences naturelles,
mmes d'affaires (cadres

cision, à classer 300 com-
mérite général. Aucun des

1969.
Writing Ability, Princeton, E.T.S., Res.

tas ne pouvait compter moins de 12 copies (4 %). 94 % de l'ensemble se sont vu attribuer 7, 8 ou 9 des 9 classements possibles. Aucune composition n'a reçu moins de cinq classements différents...

b) Mathématiques.

Le C.R.D.P. de Lyon¹ a aussi demandé à 150 professeurs de mathématiques exerçant dans une classe de 3^e de noter les copies de 3 élèves ayant à résoudre le même problème.

Max. 20	Moyenne	Marge de variation
Elève I	5,70	0,5 - 11,5
Elève II	16	11,5 - 20
Elève III	8	3,5 - 11,5

Si, pour l'élève I, par exemple, on élimine les deux notes extrêmes, il reste autant de correcteurs pour noter 2/20 que 8/20, alors qu'on pouvait attendre une notation objective.

c) Médecine.

Au niveau universitaire, les résultats ne sont pas plus sûrs.

Un même candidat de 2^e médecine a subi, en juin et en septembre 1965, une épreuve écrite comportant cinq questions simples et précises à traiter en 1 h 30.²

Les réponses (anonymes) ont été corrigées indépendamment par cinq correcteurs possédant des titres et des qualités identiques.

Résultat total sur 100	Marge de variation	Décision ³	
		Admis	Refusé
Copie de juin	47 - 78	3	2
Copie de septembre	32,5 - 73	3	2

¹ O.c.

Voir *Le Monde* du 6 septembre 1966.

³ Remarquons que les trois admis furent les mêmes dans les deux cas.

Si on considère les notes attribuées à chacune des 10 questions, on constate :

- | | | |
|--|-----|--------------------|
| 1) Ecart maximum : | 12 | (de 3 à 15 sur 20) |
| 2) Ecart pour 4 questions : | 11 | |
| 3) Plus petit écart maximum entre cinq correcteurs : | 7,5 | |

d) Divers.

A. Agazzi¹ rappelle un exemple, plus frappant encore, concernant un ensemble de branches.

Six correcteurs ont, chaque fois, noté les examens du baccalauréat (note d'échec : moins de 10 sur 20) :

	Refusés par les six correcteurs	Admis par les six correcteurs	Admis par les uns et refusés par les autres
Version latine	40 %	10 %	50 %
Composition française	21 %	9 %	70 %
Anglais	37 %	16 %	47 %
Mathématiques	44 %	20 %	36 %
Philosophie	9 %	10 %	81 %
Physique	37 %	13 %	50 %

e) Combien de correcteurs pour stabiliser la note ?

Calculer la moyenne entre deux correcteurs apporterait-il une amélioration considérable ? Non, si aucune mesure stricte n'a été prise pour accorder les notateurs. Par combien de professeurs différents faudrait-il faire évaluer le même travail pour que la note se stabilise ? Les nombres suivants, avancés par Piéron¹ n'ont certes pas de valeur absolue. Ils donnent néanmoins une idée de l'ordre de grandeur des nombres...

Composition française	78
Version latine	19
Anglais	28
Mathématiques	13
Dissertation philosophique	127
Physique	16

*révisé
noté
plusieurs*

¹ A. AGAZZI, *Les aspects pédagogiques des examens*, Strasbourg, Conseil de l'Europe, 1967, p. 119. La recherche est due à H. Laugier et D. Weinberg et portait sur le baccalauréat français.

L'histoire et la géographie de ces deux branches aussi fluctuantes.

On pourrait multiplier les notateurs. Mais, à quoi bon, si c'est facile, comme trop de remèdes.

Quelle est l'explication ?

Elle réside principalement dans la vue, des dimensions, à être jugée, et à l'absence de tout (en a pas du tout) donné.

« Dans une dissertation, les idées, leur originalité, la maîtrise du style, etc. Si l'on considère une dimension de la note, on se situe dans un espace à trois dimensions, dans cet espace, se trouvent les élèves : un élève occupant une position moins élevée que les autres ne peut dire d'un élève qu'il est supérieur à la première dimension et une position s'il est supérieur à la deuxième dimension et une position inverse. »

En outre, même si les notateurs varient en sévérité, à l'effet de halo, etc.

7. Infidélité d'un même correcteur

Un même correcteur évalue des phénomènes aussi divers que physique et mentale et l'impact du nombre de facteurs plus que la qualité de la relation de l'évaluation (si l'on veut) suivant peut être sous-estimé ou inconsciemment.

Quatorze historiens ont noté quinze compositions.

¹ F. BACHER, *La normalisation*

ouées à chacune des 10

(de 3 à 15 sur 20)

,5

e, plus frappant encore,

s, noté les examens du
e 10 sur 20) :

Admis par les six correcteurs	Admis par les uns et refusés par les autres
10 % 9 %	50 % 70 %
16 % 20 % 10 % 13 %	47 % 36 % 81 % 50 %

iser la note ?

correcteurs apporterait-il
si aucune mesure stricte
eurs. Par combien de pro-
luer le même travail pour
es suivants, avancés par
bsolute. Ils donnent néan-
eur des nombres...

...	78
...	19
...	28
...	13
...	127
...	16

mens, Strasbourg, Conseil de l'Euro-
or et D. Weinberg et portait sur le

L'histoire et la géographie ne figurent pas dans la liste ; ces deux branches aussi donnent pourtant lieu à de grandes fluctuations.

On pourrait multiplier les exemples de désaccord entre notateurs. Mais, à quoi bon se complaire dans l'accusation facile, comme trop de docimologistes l'ont fait ? L'important est d'être conscient du danger et, surtout, de chercher les remèdes.

Quelle est l'explication des divergences observées ?

Elle réside principalement dans la multiplicité des points de vue, des dimensions, selon lesquels une même épreuve peut être jugée, et à l'absence de directives précises (parfois, il n'y en a pas du tout) données pour la notation. F. Bacher¹ écrit :

« Dans une dissertation, on peut noter l'organisation des idées, leur originalité, la correction de l'expression, l'élégance du style, etc. Si l'on considère chaque aspect possible comme une dimension de la notation, le devoir d'un élève peut être situé dans un espace à n dimensions ; la difficulté vient de ce que, dans cet espace, seul un ordre partiel s'établit entre les élèves : un élève occupant une position élevée sur toutes les dimensions peut être déclaré supérieur à un élève occupant une position moins élevée sur toutes les dimensions, mais on ne peut dire d'un élève qui a une position élevée sur une première dimension et une position faible sur une deuxième dimension s'il est supérieur ou inférieur à un élève occupant des positions inverses. »

En outre, même s'ils considèrent le même aspect, les notateurs varient en sévérité, en stabilité de jugement, en résistance à l'effet de halo, etc.

7. Infidélité d'un même correcteur.

Un même correcteur est-il au moins égal à lui-même ? Non. A des phénomènes aussi évidents que les variations de santé physique et mentale et l'évolution du savoir s'ajoutent un grand nombre de facteurs plus ou moins bien définis : variation dans la qualité de la relation établie avec l'élève, dans le contexte de l'évaluation (si l'on vient de corriger un travail excellent, le suivant peut être sous-évalué,...), dans l'échelle consciemment ou inconsciemment adoptée, etc.

Quatorze historiens ont été invités à noter une deuxième fois quinze compositions, douze à dix-neuf mois après les avoir

¹ F. BACHER, *La normalisation des notes*, o.c., pp. 53-54.

notées une première fois. Toute trace de correction avait été effacée. Les professeurs accordaient non seulement des points, mais indiquaient la réussite globale ou l'échec.

Dans 92 cas sur 210, le verdict a été différent d'une fois à l'autre.¹

Il faut toutefois insister sur le fait que des résultats aussi pauvres sont dus au manque de directives rigoureuses précisant les aspects à considérer par les notateurs.

○ *Un schéma pour continuer la recherche.*

Il est utile que, de temps en temps, des recherches soient faites sur la fidélité d'un notateur par rapport à lui-même ou sur l'accord entre notateurs, ne fût-ce que pour ranimer la conscience d'un danger menaçant. C'est pourquoi nous croyons opportun de signaler l'excellent plan de recherche dû à deux des meilleurs docimologistes contemporains : F. Yates et D. Pidgeon.

- Sept groupes de 50 enfants achevant l'école primaire ont été constitués. Ils ont été soumis à deux épreuves de langue maternelle :
 - a) rédaction sur un sujet choisi parmi trois ou quatre ;
 - b) questions de compréhension de textes et questions de grammaire (usage de la langue).
- Sept examinateurs expérimentés — cinq hommes et deux femmes — ont noté les travaux.
- Chaque groupe d'enfants a d'abord subi trois examens différents, à une semaine d'intervalle ; ensuite, la première des trois épreuves a été recommencée deux fois à une semaine d'intervalle.

SCHEMA²

GROUPE	Séances d'examen				
	I	II	III	IV	V
1	Aae	Bac	Dab	Aac	Aae
2	Bbf	Cbd	Ebc	Bbf	Bbf
3	Ccg	Dce	Fcd	Ccg	Ccg
4	Dda	Edf	Gde	Dda	Dda
5	Eeb	Feg	Aef	Eeb	Eeb
6	Ffc	Gfa	Bfg	Ffc	Ffc
7	Ggd	Agb	Cga	Ggd	Ggd

¹ HARTOG et RHODES, *An Examination of Examinations*, London, McMillan, 1935, pp. 81 et p. 15.

² Il est dû à D. FINNEY, cité par Yates et Pidgeon, *Admission to Grammar Schools*, Londres, N.F.E.R., 1957, p. 99.

Sept séries de questions : A, B,

Sept examinateurs : a, b, c, d, e

Nous ne reprenons pas le détail simplement voulu montrer un exemple avec, notamment, une mesure de l'

8. **Stéréotypes et effets de halo.**

Dans la situation scolaire habituelle, chacun de ses élèves et peut donner en fonction d'un effet souhaité : imitant le travail ; là, on fait preuve de surprise pour donner un choc que l'on espère ; le maître agit délibérément, en tenant compte de l'effet de halo.

Il en va tout autrement dans les situations de halo.

a) *Stéréotypie.*

Par *stéréotypie*, nous entendons une évaluation moins accusée, qui s'installe dans une attitude.

La stéréotypie résulte d'une évaluation médiocre. Un premier travail médiocre incline à donner un jugement médiocre sera aussi ; si cela se vérifie, la tendance à donner un jugement médiocre au troisième travail s'accroît. Chez le professeur surchargé, la stéréotypie de notation se produit d'autant plus facilement.

On aurait tort de croire que la stéréotypie se limite à la notation, composition d'histoire et, en fait, à un jugement de valeur). Elle atteint aussi les jugements « types » que la dictée orthographique et la dictée à l'écrit.

Un professeur de langue maternelle qui corrige des dictées. Bientôt, il connaît par expérience habituellement le mieux et le moins bien. Il détermine la fréquence des fautes par le correcteur, on constate que les élèves sont significativement plus faibles. Dans le premier cas, le nombre de fautes contre d'erreurs ; dans le second cas, le nombre de fautes contre d'erreurs.

¹ Voir à ce propos M. ZILLIG, *Beliebte u. beliebte u. d. ges. Psychologie*, 12, 1934, p. 32, cité par R. Piper, 1967, p. 105.

trace de correction avait été
 nient non seulement des points,
 le ou l'échec.

dict a été différent d'une fois

le fait que des résultats aussi
 e directives rigoureuses préci-
 r les notateurs.

erche.

temps, des recherches soient
 par rapport à lui-même ou sur
 ce que pour ranimer la conce-
 C'est pourquoi nous croyons
 an de recherche dû à deux des
 porains : F. Yates et D. Pid-

nevant l'école primaire ont été
 à deux épreuves de langue

si parmi trois ou quatre ;
 on de textes et questions de
 ue).

és — cinq hommes et deux

oord subi trois examens diffé-
 lle ; ensuite, la première des
 cée deux fois à une semaine

IA²

s d'examen		
III	IV	V
Dab	Aac	Aae
Ebc	Bbf	Bbf
Fcd	Ccg	Ccg
gde	Dda	Dda
hef	Eeb	Eeb
ifg	Ffc	Ffc
gga	Ggd	Ggd

Examinations, London, McMillan, 1935, pp.
 Pidgeon, Admission to Grammar Schools,

Sept séries de questions : A, B, C, D, E, F, G.

Sept examinateurs : a, b, c, d, e, f, g.

Nous ne reprenons pas le détail des résultats, mais avons
 simplement voulu montrer un exemple de contrôle bien mené,
 avec, notamment, une mesure de l'apprentissage.

8. Stéréotypes et effets de halo.

Dans la situation scolaire habituelle, le maître connaît
 chacun de ses élèves et peut donc doser, nuancer ses notes
 en fonction d'un effet souhaité : ici, on encourage en suresti-
 mant le travail ; là, on fait preuve d'une sévérité exceptionnelle
 pour donner un choc que l'on espère salutaire. Dans ces cas,
 le maître agit délibérément, en toute conscience.

Il en va tout autrement dans les phénomènes de stéréotypie
 et de halo.

a) Stéréotypie.

Par *stéréotypie*, nous entendons une immuabilité plus ou
 moins accusée, qui s'installe dans le jugement porté sur l'élève.

La stéréotypie résulte d'une contamination des résultats.
 Un premier travail médiocre incline à penser que le second le
 sera aussi ; si cela se vérifie, la tendance à accorder une note
 médiocre au troisième travail s'accroît encore, et ainsi de
 suite. Chez le professeur surchargé de corrections, la défor-
 mation se produit d'autant plus facilement.

On aurait tort de croire que la stéréotypie influence uni-
 quement les évaluations à base subjective accusée (disserta-
 tion, composition d'histoire et, en général, réponses impliquant
 un jugement de valeur). Elle atteint des exercices aussi « objec-
 tifs » que la dictée orthographique. L'expérience suivante en
 témoigne.

Un professeur de langue maternelle fait régulièrement
 des dictées. Bientôt, il connaît les élèves qui réussissent
 habituellement le mieux et le moins bien cet exercice. Si l'on
 détermine la fréquence des fautes « oubliées », non perçues
 par le correcteur, on constate que les oublis en faveur des bons
 élèves sont significativement plus élevés que pour les élèves
 faibles. Dans le premier cas, le maître s'attend à ne pas ren-
 contrer d'erreurs ; dans le second, il les guette¹.

¹ Voir à ce propos M. ZILLIG, Beliebte und unbeliebte Volksschülerinnen, in Arch.
 f. d. ges. Psychologie, 12, 1934, p. 32, cité par E. HOHN, Der schlechte Schüler, Munich,
 R. Piper, 1967, p. 105.

nce professionnelle intervient, la
ve. Un cas de stéréotypie accu-
e pour les versions latines nous
nt secondaire). A titre de vérifi-
es devoirs ultérieurs successive-
a classe, par un élève de même
emier dans une autre école, puis
classique : la note n'a pas varié

genre d'expériences à d'autres
s, aussi, parfois très compétents
ent occasionnellement le travail
quelque déception ou étonnement
e de « leur » note.

maîtres incriminés de telle façon
nsabilités serait pourtant injuste.
valuation que les hommes qu'il

n caractère affectif accusé. Sou-
es d'un élève de belle allure, au
éable. Il ne faut cependant pas
professeurs qui, par anticonfor-
débraillé ou des originalités qui
r goût.

lisibilité, soit pour des raisons
peut aussi influencer le correc-
publicité savent depuis longtemps
ge exerce une influence considé-

s, dues à R. Weiss¹, montrent
raison de stéréotypie et d'effet de
provoquée.

deux rédactions faites par des
aux ont été dactylographiés sans
omisés pour correction à deux
eignant aussi en 4^e primaire.

accompagnait la distribution des
Voici deux compositions écrites
Le travail n° 1 est l'œuvre d'un

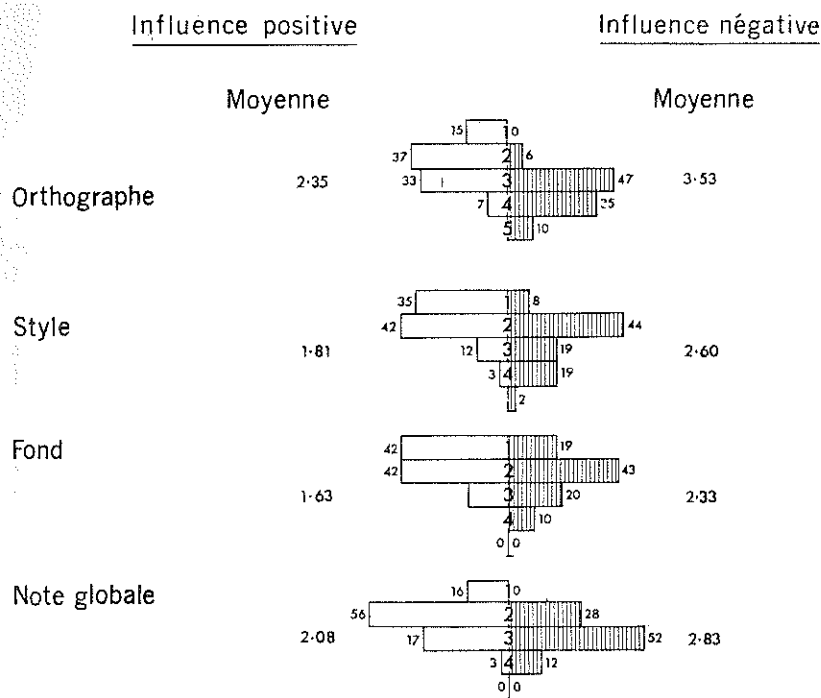
Number Marking System : An Austrian Study, in
Londres, Evans Brothers Ltd, 1969, pp. 101-107.

élève moyen qui aime lire les bandes dessinées ; son père et sa mère sont employés. Le travail n° 2 a été fait par un enfant doué ; son père est rédacteur d'un quotidien connu ». Pour le second groupe de maîtres, les rôles ont été inversés.

La correction devait se faire selon une échelle à cinq degrés : très bien (1) ; bien (2) ; moyen (3) ; suffisant (4) et insuffisant (5). L'orthographe, le style, le fond, puis l'ensemble devaient être notés séparément.

Pour les quatre aspects considérés, les notes attribuées au travail pour lequel on a créé un préjugé favorable ont été significativement supérieures aux autres. Pour l'orthographe, qui semblait le plus devoir échapper à l'effet de halo, on observe qu'au travail de l'élève présenté comme doué, 16% des correcteurs accordent la note *très bien* et aucun la note *insuffisant* ; si le même élève est présenté comme moyen, les correcteurs n'accordent aucun *très bien*, mais 11% notent *insuffisant*.

NOTATION DES COMPOSITIONS SOUS L'INFLUENCE D'UN PREJUGE FAVORABLE OU DEFAVORABLE¹

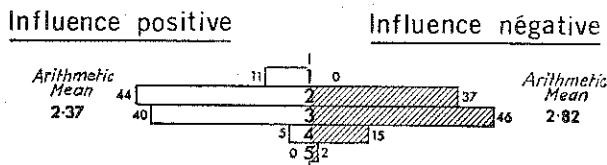


¹ Toutes les différences sont statistiquement significatives à P.01 sauf une (P.10).

Dans une expérience similaire sur des problèmes d'arithmétique, également en 4^e primaire, R. Weiss observe :

<i>Préjugé favorable</i>	<i>Préjugé défavorable</i>
11 % de très bien	Aucun très bien
5 % de suffisant	15 % de suffisant
0 % d'insuffisant	2 % d'insuffisant

PROBLEME D'ARITHMETIQUE



On voit combien un artifice relativement grossier exerce déjà une influence. Comme le souligne Weiss, la déformation est très probablement bien plus grande encore en situation réelle où le maître connaît l'apparence de l'élève, sa conduite, sa façon de s'exprimer, la profession et le statut social des parents, etc.

9. Manque de validité.

Les examens traditionnels ne sont pas seulement de pauvres moyens d'inventaire ; leur valeur pronostique apparaît tout aussi contestable dans bien des cas.

Nous avons observé, en compagnie du préfet d'un des meilleurs athénées de Belgique, que deux catégories d'élèves réussissaient brillamment dans l'enseignement universitaire : ceux qui, pendant toutes leurs études, semblent s'être joués de la difficulté et sont restés en tête de leur classe, et une partie, non négligeable, de ceux qui n'ont presque jamais enregistré d'échecs, mais se sont néanmoins maintenus juste au-dessus de la note fatidique. Dans les deux cas, les performances dans l'enseignement secondaire sont de bon augure, moins par leur nature apparente que par la facilité d'adaptation, par la plasticité intellectuelle qui y a conduit.

Pronostiquer la réussite ou l'échec en faculté sur la base du pourcentage obtenu au terme de l'enseignement secondaire est donc bien hasardeux.

10. Un instrument d'immobilisme

L'examen traditionnel se justifie comme un moyen de sélection que comme un moyen de promotion individuelle. De plus, cette justification est un coup de cas, moins en fonction de la qualification sociale.

Ainsi, l'examen devient un instrument d'immobilisme social, phénomène analysé par J.-C. Passeron à qui nous empruntons les observations suivantes¹.

a) Effets irréversibles de la certification

Une fois obtenu, un titre scolaire assure grosso modo son grade, son niveau de rémunération, son statut social.

Or, au moment de l'obtention, la compétence se trouve mesurée non pas par ce qui sera exigé du travailleur, mais par un examen analogue, préparatoire, ludique, qui ne mesure pas les examens par rapport à la tâche réelle, ce n'est pas ce que les parents veulent, mais plutôt le niveau auquel il sera payé » (p. 7).

Ces observations n'ont certes pas de valeur absolue, mais elles correspondent néanmoins à ce qui se passe, par exemple, jamais on n'attribue un poste à un élève ou au technicien qui en a acquis le titre.

Le dernier examen qu'il s'agit de passer est un individu dans une zone sociale, ce qui n'est encore aujourd'hui, fort malaisé. La fixation est en contradiction avec la réalité contemporaine d'éducation et de travail.

Que l'examen accorde plus de poids à la garantie de compétence technique que par le fait que, dans beaucoup de professions, occupent le même poste :

« l'une officiellement parce qu'elle requiert les qualifications officielles requises pour l'occuper (le diplôme et le pouvoir), l'autre effectivement parce qu'elle est la part technique de la tâche (le savoir-faire et le travail » (...). Parfois, le savoir-faire est plus important que le diplôme.

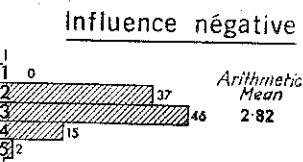
¹ J.-C. PASSERON, Sociologie des professions.

ire sur des problèmes d'arithmétique.
R. Weiss observe :

Préjugé défavorable

Aucun très bien
15 % de suffisant
2 % d'insuffisant

ARITHMETIQUE



relativement grossier exerce
souligne Weiss, la déformation
us grande encore en situation
parente de l'élève, sa conduite,
profession et le statut social des

ne sont pas seulement de pau-
r valeur pronostique apparaît
des cas.

compagnie du préfet d'un des
que deux catégories d'élèves
l'enseignement universitaire :
études, semblent s'être joués
en tête de leur classe, et une
qui n'ont presque jamais enre-
néanmoins maintenus juste au-
ans les deux cas, les perfor-
conculaire sont de bon augure,
que par la facilité d'adaptation,
y a conduit.

l'échec en faculté sur la base
de l'enseignement secondaire

10. Un instrument d'immobilisme social.

L'examen traditionnel se justifie au moins autant comme moyen de sélection que comme outil de promotion de l'éducation individuelle. De plus, cette sélection s'opère, dans beaucoup de cas, moins en fonction de la qualification technique que de la qualification sociale.

Ainsi, l'examen devient un instrument privilégié de l'immobilisme social, phénomène analysé de façon pénétrante par J.-C. Passeron à qui nous empruntons l'essentiel des considérations suivantes¹.

a) Effets irréversibles de la certification scolaire.

Une fois obtenu, un titre scolaire suit l'individu toute sa vie, lui assure *grosso modo* son grade dans la hiérarchie professionnelle, son niveau de rémunération et le pouvoir qu'il détiendra.

Or, au moment de l'obtention d'un diplôme, « la compétence se trouve mesurée non pas au travers de l'activité qui sera exigée du travailleur, mais au travers d'une sorte d'activité analogique, préparatoire, ludique et presque fictive. Ce que mesurent les examens par rapport à la demande professionnelle, ce n'est pas ce que les gens seront censés faire, mais plutôt le niveau auquel il sera socialement exigible de les payer » (p. 7).

Ces observations n'ont certes pas de valeur absolue ; elles correspondent néanmoins à une réalité incontestable. Par exemple, jamais on n'attribue le grade d'ingénieur à un ouvrier ou au technicien qui en a acquis la compétence fonctionnelle.

Le dernier examen qu'il subit avec fruit fixe donc souvent un individu dans une zone socio-professionnelle dont il est, encore aujourd'hui, fort malaisé de sortir. La précocité de cette fixation est en contradiction profonde avec toute la politique contemporaine d'éducation et donc de promotion permanente.

Que l'examen accorde plus une certification sociale qu'une garantie de compétence technique serait d'ailleurs confirmé par le fait que, dans beaucoup d'entreprises, deux personnes occupent le même poste :

« l'une officiellement parce qu'elle a tous les titres scolaires requis pour l'occuper (avec la rémunération, le prestige et le pouvoir), l'autre effectivement, parce qu'elle assume la part technique de la tâche (...), c'est-à-dire qu'elle fait « tout le travail » (...). Parfois, le second conviendrait techniquement

¹ J.-C. PASSERON, Sociologie des examens, in *Education et Gestion*, 1970, 2, 6-16.

ulaire. Mais c'est précisément ce
 tification produit par l'examen
 ur toute la vie, tend à définir ce
 nnelle d'un individu par la valeur
 des titres scolaires.»

itation appelle des restrictions.
 ompétence technique suffit rare-
 de communiquer, d'exercer des
 tions publiques, d'harmoniser le
 production, etc. sont autant de
 échappe souvent à ceux qui ne
 des choses. Mais ces qualités
 n s'en faut. Souvent, on a affaire
 ns reparler.

socialement neutres.

France, révèle que si, au cours
 nces d'accès à l'enseignement
 toutes les catégories sociales,
 ux de scolarisation s'est venti-
 tes socio-professionnelles dans
 galités antérieures. Bref, il s'agit
 e haut de la structure des iné-
 ssous est révélateur.

Catégorie	Probabilités d'accès à l'enseignement supérieur	
	1961-62	1965-66
...	1,1	2,7
...	3,4	8
...	1,3	3,4
...	9	16,2
...	24,9	35,4
supé-	38	58,7

énomène en apparence contra-
 re généralisée, de plus en plus
 et la répartition des aptitudes,
 tes les couches de la société ?

amiliale, au cours des premières
 s différences dans les mécanis-
 linguistiques et les attitudes
 donnée par les familles prédis-

pose d'autant moins à la réussite aux examens que ces familles
 appartiennent à une classe plus éloignée de la culture scolaire,
 de la culture savante » (p. 11).

Les examens « neutres » se borneraient à enregistrer les
 effets profonds de cette première éducation. Ce n'est que par-
 tiellement exact. Notre opinion est plus nuancée que celle de
 Passeron qui ne reconnaît pas assez la réalité des handicaps
 d'origine socio-culturelle auxquels on vient de faire allusion.
 Mais, comme lui, nous pensons que les examens ne sont pas
 socialement neutres.

« Les procédures de notation et les types d'épreuves utili-
 lisés prennent en compte au moins autant que les aptitudes
 techniques certains aspects gratuits de la performance, qui
 n'ont aucune importance technique, mais qui sont en revanche
 très fortement liés aux habitudes culturelles de telle classe
 sociale plutôt que de telle autre » (p. 12). Ainsi, constate Passe-
 ron, les examens français les plus prestigieux sont *moins des
 épreuves de connaissances que des épreuves de manières ou
 des exercices d'usage lettré du langage.*

La dissertation reste, dans cette perspective, le moyen
 privilégié. Rejoignant la thèse sociologique de Durkheim, de
 Max Weber, d'A. Clause, Passeron rappelle que le formalisme
 a toujours été un moyen de défense des classes privilégiées.
 Il est, en effet, frappant de constater que l'on accorde le plus
 d'importance à une épreuve, la dissertation, qui échappe le
 plus radicalement à toute notation objective, technique... Or, le
 formalisme consiste précisément à définir la culture, non par
 son contenu objectif, mais par ces impondérables que sont la
 manière, la nuance. Cette « indéfinition » permet d'exercer une
fonction de fermeture au profit d'un groupe favorisé.

La neutralité des examens scientifiques n'est d'ailleurs pas
 tellement mieux garantie, car eux aussi font souvent intervenir
 le langage, et bien peu d'examineurs restent indifférents à
 l'élégance de l'expression, pourtant étrangère au problème
 technique sur lequel porte fondamentalement l'examen.

Quant à l'examen oral, il ne fait que renforcer l'effet de
 la belle présentation, du langage châtié, de la diction élégante.

Notons, pour conclure, que la floraison des examens à ca-
 ractère sélectif paraît caractéristique des conditions culturelles
 du XIX^e siècle, en particulier du développement d'une bureau-
 cratie très hiérarchisée au service de l'économie capitaliste.
 Il est donc naturel que la nouvelle forme de civilisation vers
 laquelle nous évoluons s'accompagne d'une mise en question
 d'un type d'examen conçu pour d'autres conditions.

11. Faiblesse de beaucoup d'expériences docimologiques.

Depuis plus d'un demi-siècle, on accumule les expériences destinées à montrer les désaccords entre notateurs ou le manque de fidélité d'un même correcteur.

Le problème existe assurément. Mais on l'aggrave artificiellement, dans bien des cas, en omettant d'inviter les notateurs à s'entendre sur les aspects à considérer et l'importance relative à leur réserver. Autrement dit, on continue à enregistrer patiemment les résultats d'une politique anarchique des examens.

12. Autres critiques.

Bien d'autres critiques s'adressent encore aux examens: Le Rapport de la Commission Consultative sur les Examens dans l'Enseignement secondaire de Grande-Bretagne, *déposé dès 1911*, nous en fournit une liste. Nous l'allongeons un peu, tout en ayant l'impression de n'être guère exhaustif.

Pour l'élève :

- Il consacre trop d'énergie à reproduire les idées des autres au lieu de développer sa créativité.
- Il est récompensé pour des apprentissages souvent éphémères. (On a démontré que jusqu'à 80 % des connaissances, surtout factuelles, apprises pour l'examen, ont disparu quinze mois après.)
- Il apprend l'obéissance passive aux consignes.
- Son aptitude à s'exprimer prend parfois le pas sur le contenu. (On montrera plus tard combien les handicaps socio-culturels se marquent surtout dans le domaine verbal.)
- Il renonce souvent à exprimer un jugement personnel, pour se plier aux idées du professeur qui évaluera l'examen.
- Ses apprentissages peuvent être viciés par un esprit de compétition confinant parfois à l'esprit mercenaire.
- Il apprend à spéculer sur sa chance : dans l'examen traditionnel, le petit nombre de questions, reflet des idiosyncrasies du professeur, ouvre la voie au jeu des « tuyaux ».
- Les examens traditionnels, leur longue préparation et la période d'essoufflement qui suit, raccourcissent considérablement l'année scolaire effective.
- Les examens traditionnels empêchent le travail en groupe et exaltent, au contraire, la valeur de la performance individuelle, source d'égoïsme.

- La menace d'échec qu'ils font naître, la peur d'une sanction familiale, le succès des fraudeurs contrastent avec les élèves restés honnêtes.

- Les examens donnent une idée plus la question est complexe de conseils et d'ouvrages de

Pour le professeur :

- Il enseigne en fonction de particularités des membres d'une classe. Nous verrons qu'en bon enseignant et de l'examen doit être étranger aux objectifs des examens étrangers aux objectifs

- Il est jugulé dans sa méthode étranger aux objectifs pédagogiques.

Comment pratiquer une pédagogie personnelle, — profond d'apprentissages profonds, — caractère encyclopédique ?

- Esclave du programme, il ne trouve pas leur allure propre, et donc dignité

- Il tend à accorder trop d'importance à des connaissances utiles à l'examen

Par après, on a baptisé cette méthode par lequel les professeurs modifient de leur enseignement en fonction des exigences imposées de l'extérieur.

Des aspects éducatifs importants et l'ont souvent été.

En résumé, les examens traditionnels sont marqués de graves défauts de conception et de réalisation. L'évaluation des travaux est lourde et fastidieuse. En outre, la santé physique et mentale des élèves est compromise. Dans certains cas, les professeurs préparent d'un examen et non le contraire.

Mais il est temps de donner

CHAPITRE 2.

DEFENSE DE LA NOTE SUBJECTIVE ET DE L'EXAMEN.

Que l'on ait abusé des examens au point de vicier l'action éducative de l'école, personne ne le contestera. Que bien des concours furent surtout de sinistres loteries et duperies semble aussi hors de doute. Que des scores se virent attribuer une signification qu'ils n'avaient pas est également patent.

De là à conclure qu'examens et concours doivent être à jamais proscrits pour ne laisser subsister que des séries d'évaluations occasionnelles et les rapports qui les synthétisent, paraît utopique.

Dans la première partie, nous avons vu que les différents modes d'évaluation correspondent à des fonctions particulières. Il suffit d'en reprendre la liste pour constater qu'elles ne peuvent être toutes remplies par une seule procédure.

Par ailleurs, tous les modes d'évaluation impliquent l'établissement d'un score ou d'une note que l'on souhaite parfaitement objectifs ou, mieux, parfaitement contrôlés : entendons par là que nous reconnaissons aux maîtres le droit d'adapter beaucoup de leurs évaluations, autant en fonction de l'élève que de la matière, à condition qu'ils sachent exactement ce qu'ils font.

Avant d'envisager certains avantages souvent reconnus aux examens, une discussion de caractère négatif s'impose.

1. La mesure rigoureuse est peut-être impossible.

Rechercher les voies de la parfaite validité des examens et de la fidélité des évaluations repose sur deux postulats rappelés par J. Guillaumin¹ :

¹ J. GUILLAUMIN, L'aspect interpersonnel de la notation scolaire : de la docimologie à la doxologie, in *Bulletin de la Société A. Binet et T. Simon*, IV, 1968, p. 270.

- 1° « Que les productions de l'élève soient mesurables, quantifiables. »
- 2° « Que les différences qu'on peut constater soient mesurables, quantifiables, et pratiquées par les notateurs eux-mêmes. »

Si ces deux postulats sont acceptés, l'évaluation objective s'écroule. Or, de bons esprits ont cherché à définir de grands principes sont définis de façon précise, mais n'a guère pénétré à l'école. Ce qui est en fait, en bonne partie, à une incompatibilité fondamentale entre que de la situation d'enseignement et de la situation que ou physique des procédés d'évaluation.

Le débat n'est pas nouveau. Les progrès considérables de la psychologie contemporaine sont largement dus à l'observation, donc à la mesure. Les sciences humaines ont aujourd'hui dépassé de leurs limites et, en particulier, la psychologie générale d'utiliser des échelles de mesure parfaites. Ils savent aussi que nous ne pouvons pas mesurer rigoureusement des entités complexes. La critique, le sens de l'observation, la mesure des performances particulières supposées mesurables, est-elle considérée ?

De même que l'on n'est jamais parvenu à mesurer avec précision le rendement, tant les points de vue à considérer sont nombreux, de même on ne peut pas tenir compte de la valeur réelle d'un élève. Même si, en calcul, le résultat est faux, la démarche de la pensée est valable. varier considérablement d'un sujet à l'autre, c'est difficile à chiffrer objectivement. Qui peut prétendre que l'art est considéré comme œuvre d'art parcellaire¹. Aussi, le droit à la mesure, compte, on le comprend, est-il impossible.

Si la réponse à donner à la question de l'histoire, de sciences naturelles

¹ Signalons néanmoins qu'Ellis Page a cherché à trouver un facteur qui permet d'évaluer automatiquement le rendement, en fonction de la corrélation avec les notes traditionnelles. Dans ce cas, le chercheur n'a pas simplement constaté des imparfaites des notateurs... E. PAGE et D. PAULUS, *The Analysis of Educational Achievement*, Project 6-1318, 1968.

- 1° « Que les productions de l'élève sont par nature mesurables, quantifiables. »
- 2° « Que les différences qu'on peut trouver entre les mesures pratiquées par les notateurs sont susceptibles d'être réduites. »

Si ces deux postulats sont faux, la docimologie classique s'écroule. Or, de bons esprits constatent que, bien que ses grands principes sont définis depuis plusieurs décennies, elle n'a guère pénétré à l'école. Cet échec serait dû, pour une bonne partie, à une incompatibilité entre la nature psychologique de la situation d'enseignement et le caractère mathématique ou physique des procédés de mesure proposés.

Le débat n'est pas nouveau et l'issue est dans le compromis. Les progrès considérables de la psychologie et de la pédagogie contemporaines sont largement dus à l'objectivation de l'observation, donc à la mesure. Toutefois, les chercheurs en sciences humaines ont aujourd'hui une conscience assez claire de leurs limites et, en particulier, de l'impossibilité presque générale d'utiliser des échelles de mesure mathématiquement parfaites. Ils savent aussi que nous ne mesurons le plus souvent que des comportements isolés du tout humain. Comment évaluer rigoureusement des entités hypothétiques comme l'esprit critique, le sens de l'observation, etc., sinon à travers des performances particulières supposées représentatives de l'ensemble considéré ?

De même que l'on n'est jamais parvenu, jusqu'à présent, à mesurer avec précision le rendement de l'enseignement, tant les points de vue à considérer à des moments différents sont nombreux, de même on ne peut rendre analytiquement compte de la valeur réelle d'une performance scolaire complexe. Même si, en calcul, le résultat ne peut être que juste ou faux, la démarche de la pensée et l'effort produit peuvent varier considérablement d'un sujet à l'autre et donc être difficiles à chiffrer objectivement. Quant à la composition française, considérée comme œuvre d'art, elle échappe à l'évaluation parcellaire¹. Aussi, le droit à la subjectivité, à la réaction globale, compte, on le comprend, d'ardents défenseurs.

Si la réponse à donner à des questions de géographie, d'histoire, de sciences naturelles contraint l'élève à s'exprimer,

¹ Signalons néanmoins qu'Ellis Page a réussi à construire un programme d'ordinateur qui permet d'évaluer automatiquement les compositions de langue maternelle. La corrélation avec les notes traditionnelles est élevée. Toutefois, il faudrait savoir si, dans ce cas, le chercheur n'a pas simplement fait adopter par l'ordinateur les démarches imparfaites des notateurs...
E. PAGE et D. PAULUS, *The Analysis of Essays by Computer*, Washington, U.S. Office of H.E.W., Project 6-1318, 1968.

oralement ou par écrit, faut-il toujours faire abstraction de l'élégance du langage, de l'ordonnance de la pensée, de la rigueur du raisonnement ?

Bref, dès que l'on ne se résout plus à ramener la réussite ou l'échec à des critères rigides et souvent schématiques, les productions des étudiants échappent à la quantification automatique, impersonnelle.

Cet argument ne suffit assurément pas pour renoncer à objectiver une partie de la notation des élèves, mais une partie seulement. L'appréciation globale du maître, toute en finesse, sa sensibilité tant à la performance matérielle qu'à l'effort de dépassement et à la faiblesse humaine, doivent garder leur place.

La richesse de l'enseignement réside avant tout dans la qualité de la relation humaine qu'il crée et l'évaluation est un des aspects de cette relation. Si elle s'appauvrit au point de ne laisser subsister qu'une communication impersonnelle, le maître peut être avantageusement remplacé par une machine à enseigner.

2. Les maîtres jugent bien leurs élèves.

Si les maîtres tendent à relativiser leurs jugements par rapport au niveau moyen du groupe, leur classement à l'intérieur de celui-ci n'en possède pas moins une validité élevée.

Il suffirait donc d'appliquer des mesures permettant de rendre ces jugements comparables entre écoles (nous verrons comment au chapitre de la « modération ») pour disposer de points de repère importants.

La sûreté de jugement des maîtres s'explique par plusieurs facteurs :

1. ils fondent leur jugement sur une observation longue et continue ;
2. ils considèrent un beaucoup plus grand nombre de facteurs (notamment de personnalité) que l'examen ;
3. ils peuvent, en particulier, tenir compte de comportements exceptionnels (en telle occasion, tel élève a fait preuve d'une lucidité, d'une originalité peu communes) qui n'apparaîtront probablement pas à l'examen.

La valeur prédictive des jugements des instituteurs (ajustés pour les rendre comparables) a été bien mise en lumière par une recherche faite par la *National Foundation for Educational Research in England and Wales*¹.

¹ Voir F. YATES and D. PIDGEON, *Admission to Grammar School*, o.c., pp. 57 sq.

En 1951 et en 1952, environ Twickenham ont été examinés dans l'enseignement secondaire. Les résultats ont été étudiés deux ans.

Epreuves subies au départ :

Tests :

1. Test d'intelligence verbale (V)
2. Test standardisé de connaissances (E1) (Questions à choix multiples)
3. Test standardisé de connaissances (Calculs et problèmes) ;
4. Test d'intelligence non verbale
5. Test d'aptitude spatiale (Sp.)
6. Test d'aptitude spatiale (Sp.)
7. Test de connaissances en langues plus ouvertes que E1).

Jugement par un instituteur de

1. Prédit réussite des études (échelle à 15 degrés (M) ;
2. Classe ses élèves selon l'ordre de leur réussite à l'*Grammar School*. Ces jugements sont en fonction d'un test d'intelligence du système au chapitre de l'

Epreuves utilisées pour mesurer l'ordre de réussite à l'enseignement secondaire : ordre de réussite aux études, ajusté en fonction des objectifs d'aptitudes et de connaissances, déterminé après l'élaboration du

Le tableau ci-dessous montre la corrélation entre chaque épreuve et les résultats en cours d'études :

toujours faire abstraction de l'élé-
ance de la pensée, de la rigueur

résout plus à ramener la réussite
des et souvent schématiques, les
nappent à la quantification auto-

assurément pas pour renoncer à
ation des élèves, mais une partie
ale du maître, toute en finesse, sa
ce matérielle qu'à l'effort de dé-
maine, doivent garder leur place.

ment réside avant tout dans la
qu'il crée et l'évaluation est un
Si elle s'appauvrit au point de ne
unication impersonnelle, le maî-
t remplacé par une machine à

urs élèves.

relativer leurs jugements par rap-
pe, leur classement à l'intérieur
moins une validité élevée.

er des mesures permettant de
bles entre écoles (nous verrons
modération») pour disposer de

maîtres s'explique par plusieurs

sur une observation longue et

up plus grand nombre de fac-
nalité) que l'examen ;

enir compte de comportements
asion, tel élève a fait preuve
ité peu communes) qui n'appa-
l'examen.

ements des instituteurs (ajustés
a été bien mise en lumière par
onal Foundation for Educational
S¹.

ission to Grammar School, o.c., pp. 57 sq.

En 1951 et en 1952, environ 1 200 élèves de la région de Twickenham ont été examinés lors de leur sélection à l'entrée dans l'enseignement secondaire (*Grammar Schools*). Leurs résultats ont été étudiés deux ans, puis trois ans après.

Epreuves subies au départ :

Tests :

1. Test d'intelligence verbale (V) ;
2. Test standardisé de connaissances en langue maternelle (E1) (Questions à choix multiple) ;
3. Test standardisé de connaissances en arithmétique (A) (Calculs et problèmes) ;
4. Test d'intelligence non verbale (N/V) ;
5. Test d'aptitude spatiale (Sp. 1) (Espace à deux dimensions) ;
6. Test d'aptitude spatiale (Sp. 2) (Espace à trois dimensions) ;
7. Test de connaissances en langue maternelle (E2) (Questions plus ouvertes que E1).

Jugement par un instituteur de 6^e primaire :

1. Prédit réussite des études à la *Grammar School* sur une échelle à 15 degrés (M) ;
2. Classe ses élèves selon l'ordre de leur réussite à la *Grammar School*. Ces jugements sont ajustés par les chercheurs en fonction d'un test d'intelligence verbale (voir description du système au chapitre de la « modération ») (F).

Epreuves utilisées pour mesurer le succès dans l'enseignement secondaire : ordre de mérite fourni par le préfet des études, ajusté en fonction des résultats à une batterie de tests objectifs d'aptitudes et de connaissances, administrés immédiatement après l'élaboration du classement (S.H.A.).

Le tableau ci-dessous montre, pour le groupe de 1951, la corrélation entre chaque épreuve de prédiction et les résultats en cours d'études :

Prédicteur	Groupe de 1951 : deux ans après	Groupe de 1951 : trois ans après
Jugement de l'instituteur (F)	.821	.748
(M)	.796	.722
Test Verbal (V)	.789	.704
E2	.749	.623
A	.734	.659
E1	.729	.622
N/V	.648	.535
Sp.1	.565	.453
Sp.2	.491	.361

Le jugement ajusté des instituteurs se révèle le meilleur prédicteur. La valeur pronostique élevée du simple test verbal est aussi confirmée.

Il faut y insister : ces observations ne sont valides que dans la situation considérée : la réussite dans les *Grammar Schools*. Nous ne savons pas ce qui se passerait dans d'autres formes d'enseignement. Nous disposons néanmoins ici d'une observation strictement contrôlée dont on n'a nulle raison de croire qu'elle ne puisse s'appliquer à d'autres cas.

3. Validité limitée mais réelle des examens traditionnels.

Quelles qu'en fussent les imperfections, les examens traditionnels n'ont pas empêché notre civilisation d'atteindre un niveau scientifique, jamais égalé dans l'histoire de l'humanité. Le filtrage qu'ils ont opéré s'est donc révélé, au moins partiellement, valide.

La chose est évidente. Avec la restriction faite au chapitre précédent, on peut affirmer que les examens traditionnels ont permis d'identifier les élèves les mieux et les moins bien doués, tri capital dans la forme de civilisation que nous avons connue dans la dernière centaine d'années.

Que la procédure ait été grossière (en ce sens qu'elle a ignoré ou gaspillé des talents, au détriment des couches sociales défavorisées surtout) et injuste envers la population des élèves moyens, mis brutalement dans le même sac, n'enlève rien au fait que, dans une démocratie peu évoluée, les examens ont bien joué leur rôle.

4. S'endurcir pour la vie.

Selon les behavioristes, une elle est effectivement produite. I périodiquement, l'étudiant soit exceptionnel, à bander son éne examens, avec toutes ses impe Car la vie ne lui épargnera pas bon d'y être préparé.

Par ailleurs, la concurrence fondamentaux de notre civilisation existent et l'on peut même sout en plus là où il y a plus de candid disponibles, sinon le favoritism l'inefficacité qui l'accompagne

5. Se situer par rapport aux a

Si un classement défavorab un choc, il permet aussi de se (non dans l'absolu, mais dans Celui qui a fait de son mieux pe tions ; celui qui ne l'a pas fait être avant qu'il ne soit trop faiblesses.

En règle générale, le suc mais une certaine frustration d aiguillon.

6. Large synthèse et intégrati

Il est indéniable que les ensembles de connaissances des synthèses à l'occasion de la première fois, l'économie de parties et, éventuellement, les disciplines.

7. L'examen externe contrôle

Un examen régional ou contrôle du travail des profess

S'il ne peut plus être que programmes surabondants, f moins nécessaire que l'enseig d'études précis dans ses obj principales.

Groupe de 1951 : deux ans après	Groupe de 1951 : trois ans après
.821	.748
.796	.722
.789	.704
.749	.623
.734	.659
.729	.622
.648	.535
.565	.453
.491	.361

stituteurs se révèle le meilleur
de élevée du simple test verbal

ations ne sont valides que dans
ite dans les *Grammar Schools*.
passerait dans d'autres formes
s néanmoins ici d'une observa-
on n'a nulle raison de croire
d'autres cas.

les examens traditionnels.

perfections, les examens tradi-
otre civilisation d'atteindre un
e dans l'histoire de l'humanité.
donc révélé, au moins partielle-

la restriction faite au chapitre
les examens traditionnels ont
mieux et les moins bien doués,
sation que nous avons connue
s.

ossière (en ce sens qu'elle a
, au détrimment des couches
: injuste envers la population
ent dans le même sac, n'enlève
ratie peu évoluée, les examens

4. S'endurcir pour la vie.

Selon les behavioristes, une conduite ne s'apprend que si elle est effectivement produite. Il paraît donc souhaitable que, périodiquement, l'étudiant soit amené à faire un effort exceptionnel, à bander son énergie, à affronter l'ordalie des examens, avec toutes ses imperfections, voire ses injustices. Car la vie ne lui épargnera pas semblables épreuves, et il est bon d'y être préparé.

Par ailleurs, la concurrence, la compétition sont des traits fondamentaux de notre civilisation. L'examen et le concours existent et l'on peut même souhaiter qu'ils s'imposent de plus en plus là où il y a plus de candidats que de places ou d'emplois disponibles, sinon le favoritisme sous toutes ses formes et l'inefficacité qui l'accompagne ont libre jeu.

5. Se situer par rapport aux autres.

Si un classement défavorable à une épreuve peut donner un choc, il permet aussi de se situer par rapport aux autres (non dans l'absolu, mais dans les conditions de l'examen !). Celui qui a fait de son mieux peut ainsi mieux ajuster ses ambitions ; celui qui ne l'a pas fait a l'occasion de découvrir, peut-être avant qu'il ne soit trop tard, les conséquences de ses faiblesses.

En règle générale, le succès profite mieux que l'échec, mais une certaine frustration constitue, pour certains, un utile aiguillon.

6. Large synthèse et intégration des connaissances.

Il est indéniable que les examens portant sur de vastes ensembles de connaissances obligent l'étudiant à construire des synthèses à l'occasion desquelles il perçoit, parfois pour la première fois, l'économie de tout l'édifice, les relations entre parties et, éventuellement, les points communs entre différentes disciplines.

7. L'examen externe contrôle le professeur.

Un examen régional ou national constitue un moyen de contrôle du travail des professeurs.

S'il ne peut plus être question d'asservir les maîtres à des programmes surabondants, fixés en détail, il n'en reste pas moins nécessaire que l'enseignement se déroule selon un plan d'études précis dans ses objectifs et défini dans ses matières principales.

Un temps viendra sans doute où chaque élève travaillera selon un programme individualisé ou semi-individualisé, mais, même dans ce cas, les buts à atteindre devront être définis avant que ne s'engage l'action pédagogique.

Toujours, les parents et les autorités organisatrices auront le droit de vérifier si la mission enseignante a été accomplie.

L'existence d'un contrôle portant sur tous les éléments importants du plan d'études semble aussi une saine sauvegarde contre la laxité, le relâchement. Consacrer le temps nécessaire à l'apprentissage est louable ; musarder est condamnable. Il me souvient d'un professeur de langues qui, en deux ans, n'avait pas dépassé l'introduction du cours... Apparemment, aucun inspecteur ne s'aperçut de la chose.

Enfin, l'examen permet d'établir, au moins dans une certaine mesure, si le professeur ne concentre pas trop ses efforts sur certains élèves, au détriment des autres.

moyen pour le professeur de vérifier son enseignement

8. L'examen externe, feedback pour le professeur.

Tant pour notre santé mentale que pour la meilleure efficacité de notre action, nous avons besoin d'être informés de la validité de nos comportements.

L'examen bien conçu permet au professeur de juger de la valeur de certains aspects de son enseignement à travers les apprentissages réalisés par des élèves.

Il serait erroné d'évaluer les maîtres uniquement sur les résultats obtenus à court terme. Toutefois, seul ce genre d'évaluation semble actuellement pouvoir être effectué avec quelque rigueur scientifique. Or, tous ceux qui exercent une profession se rattachant aux sciences humaines souffrent, à vrai dire à des degrés fort divers, du manque d'information sûre sur la pertinence de leur action. C'est le souci du magistrat comme de l'assistant social, du professeur comme du prêtre.

Quand un maître peut-il dire avec certitude qu'il a bien rempli sa mission ?

*

**

En résumé, nous ne sommes partisan ni de supprimer complètement les examens, ni de renoncer entièrement à la notation subjective. Il importe d'adopter une façon de faire qui emprunte à chaque procédure ce qu'elle a de meilleur et de plus sûr. Nous allons essayer de voir comment pareil but pourrait être atteint.

oute où chaque élève travaillera
individualisé ou semi-individualisé, mais,
à atteindre devront être définis
pédagogique.

Les autorités organisatrices auront
l'enseignement a été accomplie.

portant sur tous les éléments
semble aussi une saine sauvegarde
. Consacrer le temps nécessaire
musarder est condamnable. Il
de langues qui, en deux ans,
tion du cours... Apparemment,
de la chose.

établir, au moins dans une cer-
ne concentre pas trop ses efforts
nt des autres.

pour le professeur de l'enseignement
pour le professeur.

tales que pour la meilleure effi-
avons besoin d'être informés de
s.

et au professeur de juger de la
son enseignement à travers les
élèves.

es maîtres uniquement sur les
Toutefois, seul ce genre d'éva-
voir être effectué avec quelque
ux qui exercent une profession
aines souffrent, à vrai dire à
que d'information sûre sur la
e souci du magistrat comme de
comme du prêtre.

re avec certitude qu'il a bien

partisan ni de supprimer com-
encer entièrement à la notation
ne façon de faire qui emprunte
a de meilleur et de plus sûr.
ment pareil but pourrait être

TROISIEME PARTIE

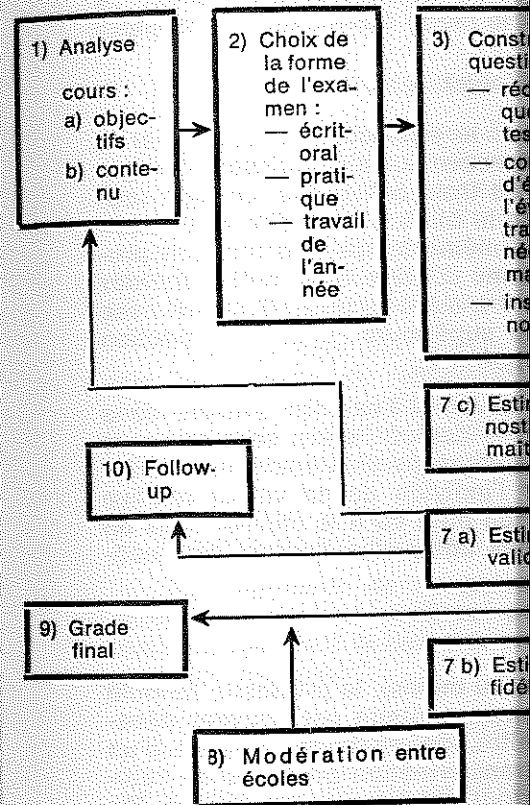
CONSTRUCTION DE L'EXAMEN

Comme les tests standardisés les maîtres devraient se construire n'est, certes, pas possible de déplication de routine, les mêmes effortinée à des centaines d'élèves. base restent les mêmes.

Les grandes phases de la con

1. Définir l'objet et les objectifs ;
2. Rédiger les questions ;
3. Standardiser la présentation, l'a
4. Etalonner ;
5. Eprouver la fidélité de l'examen
6. Etablir la validité.

Le schéma suivant montre parcourir.¹



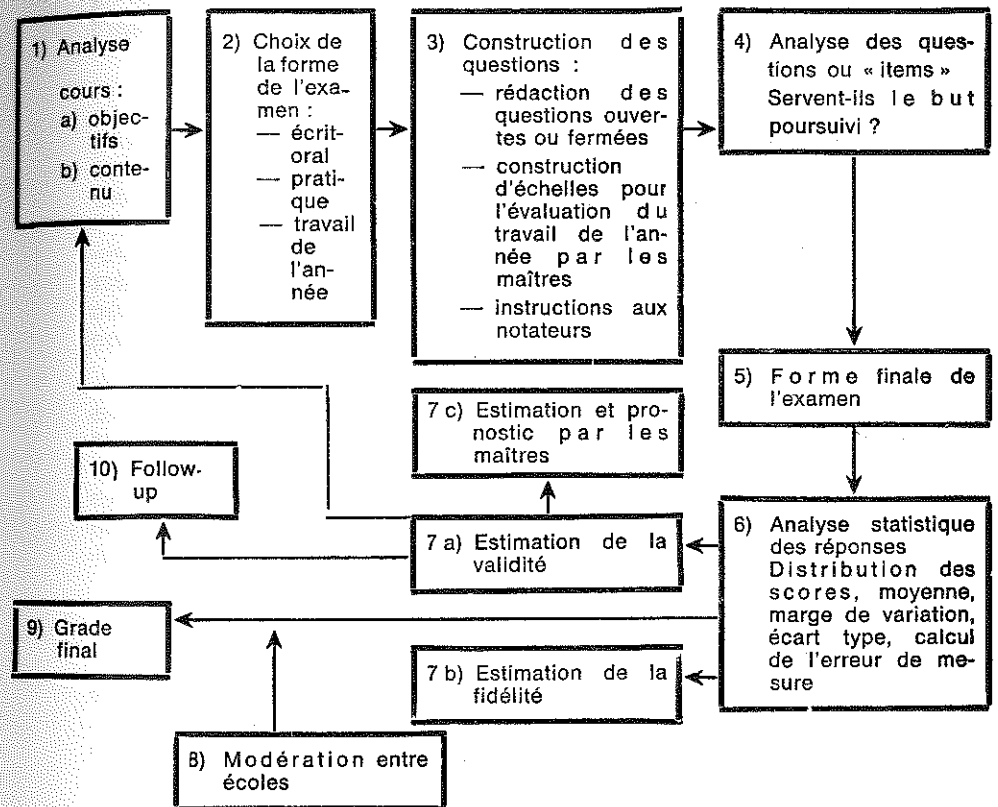
1 Adapté d'après Examinations Bulletin

Comme les tests standardisés, les examens préparés par les maîtres devraient se construire par étapes bien définies. Il n'est, certes, pas possible de déployer, pour une simple interrogation de routine, les mêmes efforts que pour une épreuve destinée à des centaines d'élèves. Pourtant, les exigences de base restent les mêmes.

Les grandes phases de la construction d'un examen sont :

1. Définir l'objet et les objectifs ;
2. Rédiger les questions ;
3. Standardiser la présentation, l'administration et la correction ;
4. Etalonner ;
5. Epruver la fidélité de l'examen ;
6. Etablir la validité.

Le schéma suivant montre le chemin que nous allons parcourir.¹



¹ Adapté d'après *Examinations Bulletin* n° 3, 1964, p. 33.

CHAPITRE 1.

L'OBJET ET LES OBJECTIFS.

Voir clairement la raison d'être de l'examen que l'on prépare et définir, avec le plus de précision possible, les types d'apprentissages à évaluer est probablement l'étape la plus décisive dans la construction d'une épreuve. Non seulement sa validité en dépend, mais aussi le mode d'évaluation à adopter !

I. L'objet.

Trois objets sont possibles : le *pronostic*, l'*inventaire* et le *diagnostique*. Il est rare qu'une épreuve scolaire puisse se ranger exclusivement dans l'une de ces trois catégories. Le plus souvent, les examens visent à faire le bilan, l'inventaire des acquisitions, mais ils remplissent en même temps une fonction pronostique à court ou à moyen terme (L'élève est-il prêt à aborder un nouveau chapitre du cours ? L'élève pourra-t-il suivre avec succès les cours de la classe supérieure ?), et une fonction diagnostique (Où l'élève achoppe-t-il ?).

Par ailleurs, il importe de savoir si l'examen s'insère dans l'action éducative quotidienne, — cas où l'élève sera comparé à lui-même et au groupe qui l'entoure (le plus souvent, sa classe), — ou s'il sert à informer l'individu de sa valeur relative dans sa communauté ou dans son pays ou à le classer dans un concours, — cas où des normes adéquates devront être utilisées.

A. Le pronostic.

A *long terme* (plus d'un an, au moins), le pronostic formulé à partir des résultats d'un examen scolaire est souvent décevant. Les changements de maître, de méthode, d'esprit d'un cours peuvent provoquer une transformation profonde de l'attitude chez l'élève. De plus, surtout chez les sujets jeunes, les intérêts manifestés connaissent d'importantes fluctuations. Il faut aussi tenir compte de l'évolution psychologique des élèves :

par exemple, les professeurs d'enseignement naissent bien l'incidence de la crise d'adap-
tation scolaire.

On a maintes fois démontré que, par
terme, les tests d'intelligence sont au
dicteurs que les tests de connaissances
sûrs que les résultats scolaires¹. Ces
études qui, surtout après l'âge de 8 ou
grande stabilité² ; elles touchent le rai-
verbal, spatial, numérique, ...

En outre, si les dominantes profon-
y compris les zones d'intérêt (et non
manifestés occasionnellement) sont ide-
succès scolaire et académique peut a-
de sûreté.

A *court terme*, le pronostic pédago-
giques :

1) Tests de maturité spécifique (reading)

Ils portent sur les formes de raison-
qui interviennent dans un apprentissage
ple : apprentissage de la lecture. La
aussi l'utilisation de ces instruments
du spécialiste. Quelques tests permet-
sont toutefois conçus pour les maîtres
d'*Inizan* pour l'apprentissage de la lect-

Dans les pays de langue franc-
que les centres de recherche aug-
leurs efforts dans le domaine de la
avons montré ailleurs combien nous
pés pour la plupart des branches et à

2) Vérification des connaissances- nécessaires aux acquisitions nouv

Des épreuves de ce genre pe-
construites par les maîtres, pour
systématiquement la liste des erreurs
gnement antérieur⁴.

1 Voir F. HOTYAT (1962), W. McCLELLAND (1962),
P. VERNON (1957), F. BACHER (1965), etc.

2 Certaines modifications peuvent toutefois en-
que le *Quotient intellectuel* n'est pas fixé une fois

3 Voir G. DE LANDSHEERE, *Les tests de*

4 Chaque notion critique doit figurer dans au-
moins deux épreuves.

par exemple, les professeurs d'enseignement secondaire connaissent bien l'incidence de la crise d'adolescence sur le rendement scolaire.

On a maintes fois démontré que, pour le pronostic à long terme, les tests d'intelligence sont au moins d'aussi bons prédicteurs que les tests de connaissances et sont beaucoup plus sûrs que les résultats scolaires¹. Ces tests explorent des aptitudes qui, surtout après l'âge de 8 ou 9 ans, connaissent une grande stabilité²; elles touchent le raisonnement, les facteurs verbal, spatial, numérique, ...

En outre, si les dominantes profondes de la personnalité, y compris les zones d'intérêt (et non des intérêts particuliers, manifestés occasionnellement) sont identifiées, la prédiction du succès scolaire et académique peut atteindre un niveau élevé de sûreté.

A court terme, le pronostic pédagogique s'établit de trois façons :

1) *Tests de maturité spécifique (readiness).*

Ils portent sur les formes de raisonnement, sur les aptitudes qui interviennent dans un apprentissage déterminé, par exemple : apprentissage de la lecture. La construction et souvent aussi l'utilisation de ces instruments complexes sont du ressort du spécialiste. Quelques tests permettant un premier dépistage sont toutefois conçus pour les maîtres, par exemple : le *Test d'Inizan* pour l'apprentissage de la lecture.

Dans les pays de langue française, il serait souhaitable que les centres de recherche augmentent considérablement leurs efforts dans le domaine de la maturité spécifique. Nous avons montré ailleurs combien nous sommes encore mal équipés pour la plupart des branches et à presque tous les niveaux³.

2) *Vérification des connaissances-clés ou notions critiques nécessaires aux acquisitions nouvelles.*

Des épreuves de ce genre peuvent être assez aisément construites par les maîtres, pour autant qu'ils aient dressé systématiquement la liste des erreurs commises lors d'un enseignement antérieur⁴.

1 Voir F. HOTYAT (1962), W. McCLELLAND (1945), E. PEEL et D. RUTTER (1951), P. VERNON (1957), F. BACHER (1965), etc.

2 Certaines modifications peuvent toutefois encore se produire. On sait aujourd'hui que le *Quotient intellectuel* n'est pas fixé une fois pour toutes.

3 Voir G. DE LANDSHEERE, *Les tests de connaissances*, Bruxelles, Editest, 1965.

4 Chaque notion critique doit figurer dans au moins trois exercices, nous y reviendrons.

Ici aussi, des fiches de recherche précisant le prérequis pour aborder les points importants de la matière rendraient aux maîtres d'inestimables services.

3) *Essai.*

Pour déterminer si un élève est prêt à aborder une matière nouvelle, on lui présente les premiers éléments de cette matière et l'on observe systématiquement les réactions. Une leçon programmée de type mixte (Skinner-Crowder) constitue souvent un heureux alliage d'enseignement et de contrôle analytique permettant un pronostic à court terme.

B. **L'inventaire** (épreuves de rendement).

Il a pour but premier de mesurer les apprentissages réalisés au cours d'une période plus ou moins longue. Sauf dans les cas de concours ou de *surveys* préparant des réformes, les inventaires remplissent aussi, à l'école, une fonction pronostique et diagnostique.

Il semble, en effet, contraire à toute attitude éducative saine qu'un maître constatant des faiblesses graves n'essaie pas d'en localiser la cause pour y remédier et ne s'inquiète pas des difficultés probables que les faiblesses annoncent pour les apprentissages ultérieurs.

Pour cette raison, notamment, il importe de distinguer soigneusement dans tout inventaire, les notions critiques, indispensables pour certains apprentissages ultérieurs, des notions marginales.

C. **Le diagnostic.**

Les épreuves de diagnostic ont pour objet de découvrir et d'expliquer les faiblesses et les habitudes défectueuses dans tous les domaines de l'apprentissage scolaire.

Sans nous aventurer dans la psychologie de l'échec (il est, dans certains cas, du ressort du psychanalyste), nous noterons, avec D. Durrell¹, que les causes les plus communes des difficultés scolaires sont :

1. Le manque d'aptitudes pour effectuer une tâche ;
2. La connaissance imparfaite des éléments de base ;
3. Un enseignement mal ajusté au niveau d'aptitudes de l'enfant et à sa vitesse d'apprentissage ;

¹ D. DURRELL, *Analysis of Reading Difficulties*, New York, Harcourt, 1955.

4. L'acquisition de mauvaises habitudes ;
5. L'inaptitude à transférer, à généraliser les techniques de solution ;
6. Le manque de vigueur, dans le cas de tant d'échecs répétés et du

Si des difficultés scolaires et la personnalité requièrent aussi, n'oublie pas toutefois que beaucoup des problèmes de l'élève proviennent de l'élève.

Dans la pratique quotidienne, l'enseignant devrait être une des préoccupations et donc occuper un temps qu'une sorte de mythe de la courtoisie fait croire, l'enseignement des notions-clés au moins, amène à une performance parfaite.

Quelques bons tests diagnostiques que les maîtres ne peuvent les ignorer, devraient construire eux-mêmes des tests simples et efficaces, spécialement

On imagine aisément l'utilité pour chacun des élèves d'une évaluation + ou par — si la réponse à des exercices portant sur un des points colorier en rouge les cases qui sont moins pour obtenir une première. Il y a quelques années, nous avons élaboré, en collaboration des élèves, un tableau de notation d'une langue étrangère.

Pour que l'épreuve puisse être faite, faut évidemment que les exercices soient précis de la matière et qu'ils soient faits de façon à ne mettre en cause

Comme l'élaboration de tests est un travail très lourd, la collaboration des élèves est souvent souhaitable.

¹ Dans une épreuve diagnostique, on a pu constater, à des endroits différents, la même notion, à des endroits différents, éventuellement joué par une distraction momentané, etc.

² Pour un exemple d'application de tests analytiques d'arithmétique (fin du primaire) voir Durrell, 1961.

recherche précisant le prérequis
tants de la matière rendraient aux
s.

ve est prêt à aborder une matière
premiers éléments de cette matière
ent les réactions. Une leçon pro-
nner-Crowder) constitue souvent
ement et de contrôle analytique
t terme.

endement).

esurer les apprentissages réalisés
i moins longue. Sauf dans les cas
éparant des réformes, les inven-
école, une fonction pronostique

raire à toute attitude éducative
des faiblesses graves n'essaie
r y remédier et ne s'inquiète pas
es faiblesses annoncent pour les

ment, il importe de distinguer
taire, les notions critiques, indis-
ntissages ultérieurs, des notions

ic ont pour objet de découvrir
les habitudes défectueuses dans
usage scolaire.

la psychologie de l'échec (il est,
u psychanalyste), nous noterons,
es les plus communes des diffi-

ffectuer une tâche ;

des éléments de base ;

é au niveau d'aptitudes de l'en-
ssage ;

difficulties, New York, Harcourt, 1955.

4. L'acquisition de mauvaises habitudes qui freinent le progrès ;
5. L'inaptitude à transférer, à découvrir le « système », à généraliser les techniques de solution ;
6. Le manque de vigueur, dans l'attaque des problèmes, résultant d'échecs répétés et du manque d'intérêt.

Si des difficultés scolaires surgissent, l'état physique et la personnalité requièrent aussi la plus grande attention. On n'oubliera pas toutefois que beaucoup de problèmes émotionnels de l'élève proviennent de l'échec au lieu d'en être la cause.

Dans la pratique quotidienne de l'école, le diagnostic fin devrait être une des préoccupations dominantes des professeurs et donc occuper un temps important. Contrairement à ce qu'une sorte de mythe de la courbe de Gauss (voir p. 180) a parfois fait croire, l'enseignement idéal serait celui qui, pour les notions-clés au moins, amènerait tous les élèves d'un groupe à une performance parfaite.

Quelques bons tests diagnostiques ont été publiés et les maîtres ne peuvent les ignorer. Mais, le plus souvent, ils pourraient construire eux-mêmes des instruments analytiques simples et efficaces, spécialement adaptés à leurs besoins.

On imagine aisément l'utilité d'un tableau d'ensemble où, pour chacun des élèves d'une classe, on voit apparaître par + ou par — si la réponse a été ou non correcte pour trois exercices portant sur un des points de la matière¹. Il suffit de colorier en rouge les cases qui contiennent deux ou trois signes moins pour obtenir une première impression de la situation². Il y a quelques années, nous avons dressé et tenu, avec la collaboration des élèves, un tableau de ce genre pour la prononciation d'une langue étrangère. Les résultats furent positifs.

Pour que l'épreuve puisse jouer son rôle diagnostique, il faut évidemment que les exercices portent sur un aspect très précis de la matière et qu'ils soient conçus, si possible, de façon à ne mettre en cause qu'une difficulté à la fois.

Comme l'élaboration de pareils contrôles devient facilement lourde, la collaboration de plusieurs professeurs est souvent souhaitable.

¹ Dans une épreuve diagnostique, on vérifie au moins trois fois la connaissance de la même notion, à des endroits différents du test, afin de réduire sérieusement le rôle éventuellement joué par une distraction momentanée, par une réponse correcte trouvée par hasard, etc.

² Pour un exemple d'application de ce système, voir : BONGRAIN, et al., *Epreuves analytiques d'arithmétique (fin du primaire)*, Morlanwelz, Institut Supérieur de Pédagogie, 1961.

II. Les objectifs.

« Il est curieux de constater que lorsqu'on demande aux personnes ayant charge d'éduquer les enfants de préciser les fins qu'elles poursuivent, on les plonge souvent dans la perplexité. »

P. Osterrieth, *Faire des adultes*, p. 9.

Affirmer que les maîtres doivent non seulement *instruire*, mais aussi *éduquer* est devenu un truisme. Encore importe-t-il de ne pas se contenter de vœux pieux, de préoccupations vagues. Certes, on enseigne ce que l'on est : notre personnalité, notre façon de penser et d'agir influencent directement, et presque malgré nous, nos élèves.

Mais, si nous voulons systématiser notre action, une définition précise des buts poursuivis devient nécessaire.

Eduquer, c'est mettre en œuvre les moyens propres à développer l'intelligence et la personnalité dans le sens voulu par le milieu culturel ; sans oublier que la santé physique exige aussi l'attention de l'éducateur.

Instruire signifie mettre quelqu'un en possession de connaissances nouvelles.

Les *objectifs généraux* relèvent de l'éducation, les *objectifs spéciaux*, de l'instruction. Mais les deux sont inséparables ; du moins, ils devraient l'être.

Evidemment, c'est *avant le début de l'année scolaire* que le professeur doit s'interroger sur les objectifs à atteindre (sinon il agit en aveugle) et *donc aussi définir la matière des examens*.

Certes, les plans d'études et les programmes scolaires officiels apportent une partie des réponses. Mais il est clair qu'elles n'apparaîtront qu'après une étude minutieuse et une longue méditation de ces documents de base. Il n'est d'ailleurs pas rare, surtout dans l'enseignement secondaire, que chaque branche fasse l'objet d'un programme séparé et que seuls les objectifs privilégiés soient explicitement traités. Il appartient alors au professeur d'ajouter les objectifs plus généraux qu'il poursuit en commun avec ses collègues des autres disciplines. Si le maître n'a pas une conscience vive des objectifs et ne les a pas faits siens, il a bien peu de chances de les atteindre.

Dresser un tableau des objectifs à poursuivre aide beaucoup le professeur au moment où il prépare sa matière de l'année. Ce tableau (voir exemple p. 66) contiendra autant de colonnes que d'objectifs généraux. Pour chaque chapitre ou

chaque étape du programme, on souhaite atteindre et l'importance (échelle à 3 ou 5 degrés).

L'appréciation de l'importance du programme échappe à la question d'une matière de base qu'il faut en aborder d'autres ? s'agit-il de domaines indispensables pour l'exercice de l'insertion dans la vie ?

Si subjective l'évaluation doit être, l'effort de réflexion qui précède presque toujours, une clarification

A. LES OBJECTIFS

Au fond, l'objectif de l'éducation de l'adulte. P.A. Osterrieth a consacré un thème ¹.

Plus spécialement, les maîtres ne pas nuire à la santé physique de l'élève autant que possible, aider à poursuivre systématiquement ce que nous allons essayer de préciser

Séparer les domaines cognitifs, affectifs, sociaux ; fait ou irrité ; aimer ou rejeter le théorique. Peut-on penser sans penser ? On le sait bien, qui répond à une stimulation, une analyse subtile devrait reconnaître les composantes cognitive, affective, sociale.

Notre propos est bien plus lié à la recherche de points de convergence entre enseignement et examens. Nous ne pouvons arbitrairement le domaine de la recherche donner à ce dernier mot une

1. Les objectifs cognitifs.

On doit à deux chercheurs, Guilford ³, des classifications hiérarchiques d'objectifs qui, malgré certaines faiblesses

1 P.A. OSTERRIETH, *Faire des adultes*, premier chapitre : « Le but de l'éducation ».

2 B. BLOOM et al., *Taxonomie des objectifs*, traduit par M. Lavallée, Montréal, Education Nationale.

J.-P. GUILFORD, *Modèle tridimensionnel des graphies* ont été publiées sur ce sujet par Guilford, California, Los Angeles).

Voir spécialement la critique de J. CA

objectifs.

est curieux de constater que lorsqu'on demande aux personnes ayant charge d'éduquer des enfants de préciser les fins qu'elles poursuivent, on les plonge souvent dans la perplexité.»

P. Osterrieth, *Faire des adultes*, p. 9.

doivent non seulement instruire, mais aussi éveiller l'intérêt, combattre le conformisme. Encore importe-t-il de ne pas tomber dans le piège des préoccupations superficielles que l'on est : notre personnalité, nos valeurs, agir influencent directement, et

organiser notre action, une définition plus précise devient nécessaire.

utiliser les moyens propres à développer l'autonomie dans le sens voulu par la pédagogie, que la santé physique exige

lorsqu'un élève est en possession de connaissances

de l'éducation, les objectifs et les méthodes sont inséparables ;

au début de l'année scolaire que l'on définit les objectifs à atteindre (sinon les objectifs de l'année) et à définir la matière des examens.

et les programmes scolaires sont des réponses. Mais il est clair que l'élaboration d'une étude minutieuse et une analyse de base. Il n'est d'ailleurs pas étonnant que, dans un enseignement secondaire, que chaque élève soit traité comme un individu et que seuls les objectifs généraux soient traités. Il appartient à l'enseignant de définir les objectifs plus généraux qu'il définit dans les autres disciplines. On ne vise pas les objectifs et ne les définit pas de chances de les atteindre. Les objectifs à poursuivre aide beaucoup à préparer sa matière de l'année (p. 66) contiendra autant de pages que l'année. Pour chaque chapitre ou

chaque étape du programme, on indique les objectifs que l'on souhaite atteindre et l'importance relative à leur réserver (échelle à 3 ou 5 degrés).

L'appréciation de l'importance relative des différents points du programme échappe à la quantification rigoureuse : s'agit-il d'une matière de base qu'il faut nécessairement connaître pour en aborder d'autres ? s'agit-il de connaissances ou de capacités indispensables pour l'exercice d'une profession ou pour l'insertion dans la vie ?

Si subjective l'évaluation de l'importance puisse-t-elle parfois être, l'effort de réflexion qu'elle exige entraîne néanmoins, presque toujours, une clarification utile.

A. LES OBJECTIFS GENERAUX.

Au fond, l'objectif de l'enseignement est unique : faire un adulte. P.A. Osterrieth a consacré des pages pénétrantes à ce thème¹.

Plus spécialement, les maîtres doivent, d'une part, au moins ne pas nuire à la santé physique et mentale de leurs élèves et, autant que possible, aider à la développer, et, d'autre part, poursuivre systématiquement des objectifs cognitifs et affectifs que nous allons essayer de préciser.

Séparer les domaines cognitif (penser), affectif (être satisfait ou irrité ; aimer ou rejeter) et conatif (vouloir, désirer) est théorique. Peut-on penser sans éprouver de sentiment, agir sans penser ? On le sait bien, c'est toujours l'organisme total qui répond à une stimulation et, pour chaque comportement, une analyse subtile devrait révéler l'importance relative des composantes cognitive, affective et psycho-motrice.

Notre propos est bien plus modeste. Nous sommes actuellement à la recherche de points de repère qui serviront à jaloner l'enseignement et les examens. C'est pourquoi nous allons couper arbitrairement le domaine cognitif du domaine affectif et donner à ce dernier mot une acception très large.

1. Les objectifs cognitifs.

On doit à deux chercheurs américains, B. Bloom² et J.-P. Guilford³, des classifications hiérarchisées des objectifs cognitifs qui, malgré certaines faiblesses⁴, se révèlent d'utiles outils.

1 P.A. OSTERRIETH, *Faire des adultes*, Bruxelles, Dessart, 1964. Voir spécialement le premier chapitre : « Le but de l'éducation ».

2 B. BLOOM et al., *Taxonomie des objectifs pédagogiques. I. Domaine cognitif*, traduit par M. Lavallée, Montréal, Education Nouvelle, 1969.

3 J.-P. GUILFORD, *Modèle tridimensionnel de l'intellect*. Une quarantaine de monographies ont été publiées sur ce sujet par Guilford et son équipe (University of South California, Los Angeles).

Voir spécialement la critique de J. CARROLL.

des articulations ; un exemple
sé page 68.

tre accepté dans le sens retenu
e¹. Par ailleurs, même si con-
paraissent comme les compor-
plus nobles, cela n'implique en
disparaître de notre enseigne-
, il faut avoir franchi chaque
supérieur. La taxonomie a été
ne pas oublier certaines étapes
lement le niveau de son ensei-

es : appellations, faits, dates,

es données particulières (sans
tions, classifications, critères,

s : principes, lois, théories, ...

sous-catégories est plus quan-
par cœur, sans plus, les dates
e ou tel exposé de la philoso-
les efforts de mémorisation dif-
te, dans les deux cas, très bas.
men, du primaire à l'enseigne-
niveau (Que savez-vous de ... ?
Comment prépare-t-on le... ?).

analyse plus ou moins fine. Ne faudrait-il
Bloom et ses collaborateurs ont rétréci le
difficulté. Ils indiquent d'ailleurs à plusieurs
entre certains échelons et analysent, avec
n de la *Taxonomie*. Notre but n'est pas d'en
er l'attention sur l'instrument.

2. COMPRENDRE.

Il s'agit ici de la compréhension au niveau le plus bas. Par une formulation nouvelle du donné, l'individu montre qu'il a dépassé le psittacisme, que le message a, pour lui, une signification. Toutefois, il n'est pas encore question d'appliquer, donc de percevoir les rapports du donné avec d'autres matériaux, d'autres situations.

Bloom distingue deux échelons :

2.1. Traduire, transposer.

Le contenu de la communication est conservé sans que son ordre soit modifié, mais la forme est changée.

Exemples : Paraphraser un récit, une proposition : « Une *taxonomie* est une *classification* ».

Exprimer verbalement des symboles mathématiques :

$A > B$ signifie que A est plus grand que B.

2.2. Interpréter.

Expliquer ou résumer une communication.

L'interprétation implique un nouvel arrangement, une nouvelle vue du matériel. Elle suppose donc la capacité de reconnaître et de saisir les idées maîtresses d'une communication et de comprendre les rapports existant entre elles. « A cet égard, l'interprétation devient synonyme de l'analyse et possède certaines caractéristiques de l'évaluation¹ ».

Exemple : Interpréter des données représentées sous forme de tableaux ou de graphiques en en tirant des déductions tenant compte des relations entre données, ou de leur signification d'ensemble.

Aller au-delà des données et des renseignements fournis : extension des tendances, généralisation.

2.3. Extrapoler.

3. APPLIQUER.

L'application suppose que le sujet distingue les traits communs à deux situations, à deux problèmes ; une abstraction s'est donc produite.

Exemple : Emploi de procédés expérimentaux pour résoudre des problèmes de travaux ménagers.

¹ B. BLOOM et al., o.c., p. 104.

4. ANALYSER.

4.1. Rechercher des éléments.

Exemple : Distinguer les faits des hypothèses dans une communication.

4.2. Rechercher des relations.

Exemple : Les hypothèses sont-elles logiques par rapport aux renseignements dont on dispose ?

4.3. Rechercher des principes d'organisation.

Exemple : Identifier les techniques de propagande utilisées dans des tracts.

5. SYNTHETISER.

5.1. Produire une œuvre personnelle.

Exemple : Narration captivante d'une expérience vécue.

5.2. Elaborer un plan d'action répondant aux exigences fixées.

5.3. Dériver un ensemble de relations abstraites. Induire une règle.

6. EVALUER.

Des jugements qualitatifs ou quantitatifs établissent dans quelle mesure le matériel et les méthodes répondent aux critères (internes ou externes).

Exemples : Déceler les sophismes dans une discussion. Apprécier un travail par comparaison à un modèle.

b) Le modèle de J.-P. Guilford.

Moins utilisé, jusqu'à présent, que la taxonomie de Bloom pour la construction des examens, le modèle de J.-P. Guilford offre peut-être de plus grandes possibilités encore, en raison de sa rigueur.

Voici d'abord comment J.-P. Guilford et R. Marrifels définissent les trois dimensions de l'intellect et leurs composantes¹ :

¹ Voir les *Reports from the Psychological Laboratory of the University of Southern California*, publiés depuis juin 1950.

LES OPERATIONS.

Ce sont les activités ou les opérations ; c'est ce que fait l'organisme informationnelle, à partir de :

1. Cognition.

Conscience, appréhension, connaissance, compréhension, formes.

2. Mémoire.

Rétention d'informations.

3. Production convergente.

Génération d'informations acceptées, à partir d'un donné ; les données sont respectés.

4. Production divergente.

Génération d'informations variées. Originalité, créativité.

5. Evaluation.

Prise de décisions ou formulations. L'exactitude, l'adéquation, l'originalité, des critères, à des idéaux, à des buts.

LES CONTENUS.

1. Figuratifs.

Information dans sa forme concrète. Images.

Un minimum d'organisation. Intelligence pratique.

2. Symboliques.

Informations sous forme de symboles. Information par et en eux-mêmes. Intelligence théorique.

Intelligence théorique.

3. Sémantiques.

Informations sous forme de mots. Intelligence verbale.

Intelligence verbale.

LES OPERATIONS.

Ce sont les activités ou les processus intellectuels principaux ; c'est ce que fait l'organisme à partir de la matière première informationnelle, à partir de ce qu'il discrimine.

1. Cognition.

Conscience, appréhension, découverte ou redécouverte, reconnaissance, compréhension d'informations sous diverses formes.

2. Mémoire.

Rétention d'informations.

3. Production convergente.

Génération d'informations uniques, conventionnellement acceptées, à partir d'un donné. L'usage, la coutume, la règle sont respectés.

4. Production divergente.

Génération d'informations variées à partir d'un même donné. Originalité, créativité.

5. Evaluation.

Prise de décisions ou formulation de jugements concernant l'exactitude, l'adéquation, la désirabilité, ... conformément à des critères, à des idéaux, à des objectifs adoptés.

LES CONTENUS.

1. Figuratifs.

Information dans sa forme concrète, perçue ou rappelée en images.

Un minimum d'organisation, de structuration est nécessaire. Intelligence pratique.

2. Symboliques.

Informations sous forme de signes dépourvus de signification par et en eux-mêmes : lettres, nombres, notes de musique.

Intelligence théorique.

3. Sémantiques.

Informations sous forme de significations attachées à des mots.

Intelligence verbale.

4. Comportementaux.

Informations, essentiellement non verbales, intervenant dans les interactions humaines, où la perception d'attitudes, de besoins, de désirs, d'intentions, de pensées d'autrui et de soi-même jouent un rôle.

Intelligence sociale.

LES PRODUITS.

Ce sont les résultats du traitement des informations par l'organisme.

1. Unités.

Portions d'information relativement isolées ou circonscrites.

2. Classes.

Unités groupées en raison de leurs propriétés communes.

3. Relations.

Connexions reconnues entre des unités.

4. Systèmes.

Groupements d'unités organisées ou structurées ; complexes de parties se trouvant en interrelation ou en interaction.

5. Transformations.

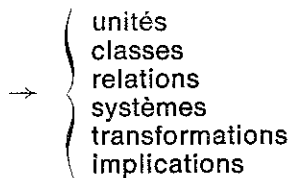
Changements apportés dans des informations ou dans leur utilisation.

6. Implications.

Extrapolation d'informations : prédiction, conséquences, antécédents.

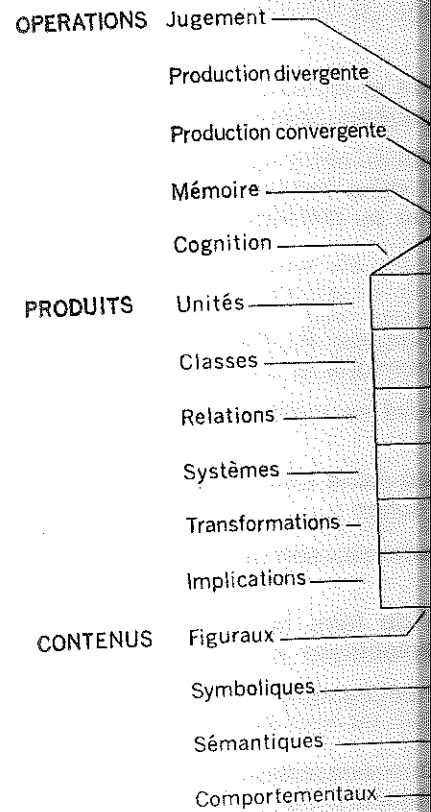
Chacune des composantes des trois dimensions se combine avec toutes les autres. Ainsi :

Mémoire des



soit 24 combinaisons.

Comme on distingue cinq types de produits en tout, cent vingt combinaisons.



L'expérience montre que les combinaisons en termes de produits sont nombreuses.

Dans un premier temps, les enseignants retiendront surtout l'attention sur la scolarité, le maître devra organiser toutes les activités.

non verbales, intervenant dans
la perception d'attitudes, de
s, de pensées d'autrui et de

itement des informations par

ment isolées ou circonscrites.

leurs propriétés communes.

es unités.

sées ou structurées; com-
n interrelation ou en interac-

es informations ou dans leur

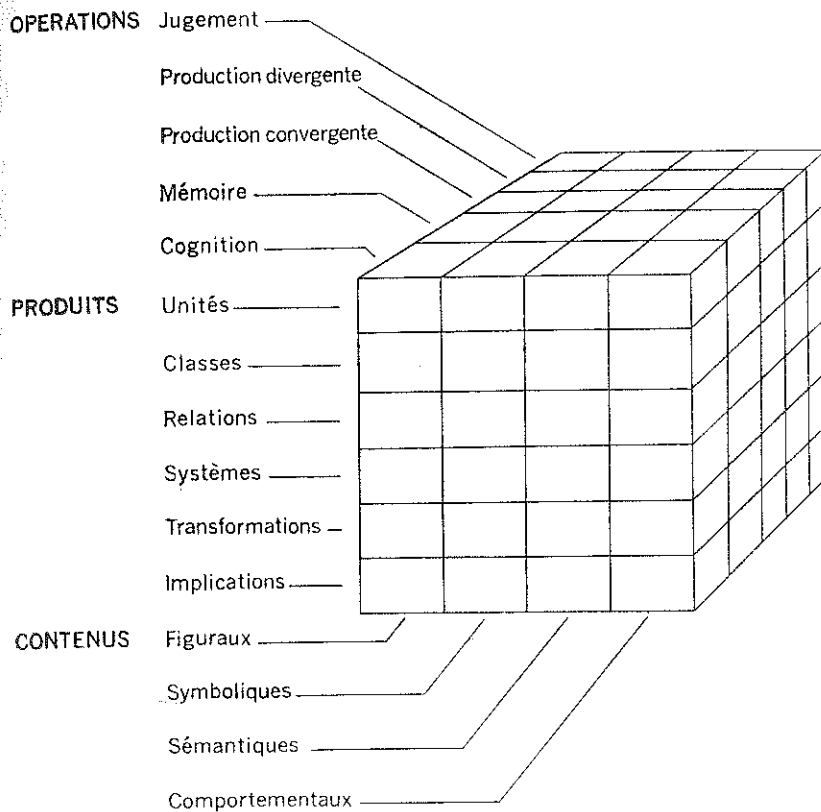
prédiction, conséquences,

es trois dimensions se com-

→ {
figuratifs
symboliques
sémantiques
comportementaux

soit 24 combinaisons.

Comme on distingue cinq types d'opérations, il existe donc, en tout, cent vingt combinaisons dans le modèle.



L'expérience montre qu'il est difficile de traduire toutes les combinaisons en termes utilisables pour l'enseignement.

Dans un premier temps, les quatre types d'opérations retiendront surtout l'attention. Idéalement, et dès le début de la scolarité, le maître devrait s'efforcer de les introduire dans toutes les activités.

1. La restitution de mémoire témoigne de la fixation de la notion.
2. La production convergente témoigne de la capacité d'appliquer les acquisitions en conformité avec les règles, les usages.
3. La production divergente témoigne de la capacité de découvrir des solutions ou des usages nouveaux.
4. Enfin, l'évaluation témoigne de la capacité de critiquer, de porter un jugement en fonction de critères bien définis et d'agir en conséquence.

Trop souvent, l'activité scolaire et les examens qui la sanctionnent se limitent aux deux premières opérations.

S'efforcer de garder un équilibre entre les quatre catégories des contenus importe autant dans l'enseignement général que dans l'enseignement technique ou artistique.

2. Domaine affectif.

En schématisant l'analyse fine de P. Osterrieth¹, nous dirons qu'un individu accède pleinement à la condition d'adulte :

- 1° a) Si son comportement a trouvé sa logique, sa cohérence et échappe à la versatilité ;
- b) S'il a, par ailleurs, acquis une saine tolérance au changement, à la contradiction, à l'échec ;
- 2° a) S'il a conquis son indépendance, son autonomie intellectuelle et affective ;
- b) S'il est cependant capable de faire le don de soi, de rester fidèle à ses engagements et à ses sentiments.

L'éducation est un long acheminement vers cet ultime équilibre.

Nous proposons ci-dessous une adaptation et une interprétation de la taxonomie des objectifs de D. Kratochvíl et B. Bloom². On obtient de nouveau cinq échelons, cinq étapes qui acheminent du comportement le plus passif au plus actif.

1 P. OSTERRIETH, *Faire des adultes*, Bruxelles, Dessart, 1964.

2 D.R. KRATOWOHL, B.S. BLOOM, B. MASIA, *Taxonomie des objectifs pédagogiques, Domaine affectif*, Montréal, Education nouvelle, 1970.

I. L'individu répond à un

1. Est simplement réceptif.

C'est une sorte d'état affectif où la beauté ou la laideur, les sens, se reflètent peu comme un miroir qui ne réagit pas.

Ce comportement est d'ailleurs une simple cognition qui précède l'émotion. Un certain éveil de l'attention est observé, on écoute, on regarde, on entend, on voit, on sent, on goûte, on touche, on entend parler les autres.

2. Reçoit et réagit.

L'individu réagit nettement à l'émotion, manifestant du plaisir, par la parole, par le geste. À ce stade, on n'observe pas encore de choix délibéré.

Pour le professeur de littérature, les élèves n'ont pas encore le goût du choix personnel, où leur sensibilité est raffinée pour leur permettre de choisir, mais où, mis en présence de la grandeur, ils en sentent la grandeur.

3. Reçoit et réagit en acceptant.

Maintenant, l'individu sait choisir, il est en condition d'être mis en présence de la grandeur, il s'engage.

II. L'individu p

4. Essaie spontanément de co

L'individu éprouve assez de confiance pour agir sans y être invité, assez de confiance pour une initiative sentimentale ou, encore, assez de sens des valeurs pour se consacrer à la religion.

5. Agit selon ses options.

C'est le stade psychologique où l'individu agit selon ses options. P. Osterrieth.

témoigne de la fixation de la

témoigne de la capacité d'appli-
conformité avec les règles, les

moigne de la capacité de décou-
es nouveaux.

de la capacité de critiquer, de
ion de critères bien définis et

ire et les examens qui la sanc-
mières opérations.

quilibre entre les quatre catégo-
t dans l'enseignement général
ue ou artistique.

ine de P. Osterrieth¹, nous
ement à la condition d'adulte :
uvé sa logique, sa cohérence

ne saine tolérance au change-
échec ;

ndance, son autonomie intel-

de faire le don de soi, de
ements et à ses sentiments.

neminement vers cet ultime

ne adaptation et une inter-
bjectifs de D. Kratochvíl et
cinq échelons, cinq étapes
le plus passif au plus actif.

¹, Dessart, 1964.

Taxonomie des objectifs pédagogiques,
1970.

I. L'individu répond à une stimulation extérieure.

1. Est simplement réceptif.

C'est une sorte d'état affectif amorphe où le sujet perçoit la beauté ou la laideur, les sentiments divers, sans réagir, un peu comme un miroir qui ne renverrait pas l'image.

Ce comportement est d'ailleurs difficile à distinguer de la simple cognition qui précède la mise en mémoire. Seul, un certain éveil de l'attention est observable. Exemple : Ecoute la musique, écoute parler les autres.

2. Reçoit et réagit.

L'individu réagit nettement soit en obéissant, soit en manifestant du plaisir, par la parole, par le geste ou l'attitude. A ce stade, on n'observe pas encore de rejet explicite qui témoignerait d'un choix délibéré.

Pour le professeur de littérature, c'est le moment où les élèves n'ont pas encore le goût assez formé pour faire un choix personnel, où leur sensibilité n'est pas encore assez raffinée pour leur permettre de partir seuls à la découverte, mais où, mis en présence de belles œuvres, ils commencent à en sentir la grandeur.

3. Reçoit et réagit en acceptant ou en refusant.

Maintenant, l'individu sait ce qu'il veut ou ce qu'il aime, à condition d'être mis en présence des personnes ou des choses ; il s'engage.

II. L'individu prend l'initiative.

4. Essaie spontanément de comprendre, de juger, de ressentir.

L'individu éprouve assez d'intérêt, de curiosité, pour s'instruire sans y être invité, assez de sensibilité pour prendre une initiative sentimentale ou, encore, il a suffisamment découvert le sens des valeurs pour se choisir une philosophie ou une religion.

5. Agit selon ses options.

C'est le stade psychologiquement adulte, comme l'a défini P. Osterrieth.

Par exemple, l'individu vit en fonction de ses options morales, sentimentales, esthétiques, mais il est aussi capable de changer de conduite à la lumière de preuves, d'arguments convaincants.

Cette ultime étape de l'ascension affective correspond à l'évaluation dans le domaine cognitif.

B. LES OBJECTIFS SPECIAUX.

Théoriquement, toutes les matières, tous les points des programmes scolaires offrent l'occasion de se rapprocher des objectifs généraux et de vérifier s'ils ont été atteints.

Trois remarques s'imposent toutefois :

- 1° La pratique montre qu'une même épreuve ne peut porter que sur un nombre limité d'objectifs ;
- 2° Certaines matières se prêtent mieux que d'autres à la poursuite de certains objectifs ;
- 3° Le grave problème du transfert des apprentissages continue à se poser avec acuité. Par exemple, cultiver la divergence à l'occasion des activités artistiques ne garantit en rien que ce trait se manifesterait dans les domaines scientifiques ou dans la vie pratique.

Quoi qu'il en soit, la première démarche visant à l'identification des objectifs spéciaux reste fondamentalement toujours la même et consiste en l'élaboration d'un tableau à double entrée. En haut, on porte les objectifs généraux ; sur le côté, on inscrit les matières du cours. Chaque intersection de colonne et de rangée indique, en principe, un objectif spécial.

Exemple : CHIMIE ¹.

	Connaissance des faits, méthodes et techniques	Application	Évaluation
Équilibres ioniques ¹			
1. Généralités			
1. Degré d'ionisation ou fraction ionisée			
2. Mesures du degré d'ionisation			
3. Application de la loi d'action des masses aux équilibres ioniques			
4. Loi de la dilution d'Ostwald			
2. Produit de solubilité			
1. Définition			
2. Calcul du produit de solubilité en fonction de la solubilité			

¹ Extrait de : BERGER et DIGHAYE, *Chimie IV*, Liège, Sciences et Lettres, 1967.

3. Diminution de la solubilité d'un électrolyte peu soluble
 4. Précipitation d'un électrolyte par adjonction, à sa solution saturée, d'un électrolyte apportant un ion commun
 5. Dissolution des précipités
- Appendice : Application du produit de solubilité à la précipitation des sulfures
3. Produit ionique de l'eau et pH
 1. Produit ionique de l'eau
 2. Le pH
 - a La notation pH
 - b pH de l'eau pure
 - c L'échelle des pH
 - d Détermination du pH
 - e Les indicateurs colorés
 - f Méthodes expérimentales de détermination du pH
 - g Distinction entre électrolytes (acides et bases) forts et faibles, par détermination du pH de solutions diluées de concentrations connues
 - h Calcul du pH des solutions d'acides et de bases à partir de leurs concentrations molaires et — éventuellement — leurs constantes d'ionisation
 4. Solutions tamponnées
 1. Introduction expérimentale
 2. Calcul du pH d'un mélange tampon
 5. Hydrolyse
 1. Généralités
 2. Cas de sels dérivant d'un acide fort et d'une base forte
 3. Cas des sels dérivant d'un acide fort et d'une base faible
 4. Cas des sels dérivant d'un acide faible et d'une base forte
 5. Cas des sels dérivant d'un acide faible et d'une base faible
 6. Facteurs influençant l'hydrolyse et des bases
 7. Les méthodes d'analyse quantitative par voie chimique
 1. Méthodes gravimétriques
 2. Méthodes volumétriques
 - a Acide fort - base forte
 - b Acide faible - base forte
 - c Base faible - acide fort

fonction de ses options morales mais il est aussi capable de mener de preuves, d'arguments

dimension affective correspond à un objectif.

OBJECTIFS SPECIAUX.

matières, tous les points des programmes ont été atteints.

Outrefois :

même épreuve ne peut porter sur des objectifs ;

mieux que d'autres à la poursuite

de la poursuite des apprentissages continue. Par exemple, cultiver la divergence des méthodes ne garantit en rien que dans les domaines scientifiques

la démarche visant à l'identification fondamentale toujours d'un tableau à double entrée d'objectifs généraux ; sur le côté, on a une intersection de colonne et d'objectif, un objectif spécial.

CHIMIE 1.

Connaissance des faits, méthodes et techniques	Application	Evaluation

	Connaissance des faits, méthodes et techniques	Application	Evaluation
3. Diminution de la solubilité d'un électrolyte peu soluble			
4. Précipitation d'un électrolyte par adjonction, à sa solution saturée, d'un électrolyte apportant un ion commun			
5. Dissolution des précipités			
Appendice : Application du produit de solubilité à la précipitation des sulfures			
3. Produit ionique de l'eau et pH			
1. Produit ionique de l'eau			
2. Le pH			
a La notation pH			
b pH de l'eau pure			
c L'échelle des pH			
d Détermination du pH			
e Les indicateurs colorés			
f Méthodes expérimentales de détermination du pH			
g Distinction entre électrolytes (acides et bases) forts et faibles, par détermination du pH de solutions diluées de concentrations connues			
h Calcul du pH des solutions d'acides et de bases à partir de leurs concentrations molaires et — éventuellement — de leurs constantes d'ionisation			
4. Solutions tamponnées			
1. Introduction expérimentale			
2. Calcul du pH d'un mélange-tampon			
5. Hydrolyse			
1. Généralités			
2. Cas de sels dérivant d'un acide fort et d'une base forte			
3. Cas des sels dérivant d'un acide fort et d'une base faible			
4. Cas des sels dérivant d'un acide faible et d'une base forte			
5. Cas des sels dérivant d'un acide faible et d'une base faible			
6. Facteurs influençant l'hydrolyse			
6. Neutralisation mutuelle des acides et des bases			
1. Chaleur de neutralisation			
2. Courbes de neutralisation			
a Acide fort - base forte			
b Acide faible - base forte			
c Base faible - acide fort			
7. Les méthodes d'analyse quantitative par voie chimique			
1. Méthodes gravimétriques			
2. Méthodes volumétriques			
Acides - bases			
Choix des indicateurs			
Titrages rédox			

La matière de chimie figurant dans le tableau ci-dessus constitue un des quatre chapitres d'un cours destiné à l'année supérieure de l'enseignement secondaire. Parmi bien d'autres objectifs généraux possibles, nous n'en avons retenu que trois.

Un simple coup d'œil sur l'ensemble révèle que, même dans ces conditions très simplifiées, le professeur se trouve devant un choix difficile. Sur quels points va-t-il faire porter l'examen pour obtenir un échantillonnage suffisant de la matière ? Quelle importance relative va-t-il réserver aux matières retenues et aux divers objectifs spéciaux qui y correspondent ? En quoi consiste exactement l'évaluation dans un cours comme celui-ci ?

On le voit, le problème est loin d'être simple. Il dépasse presque toujours la compétence d'un seul homme et débouche sur des questions restées sans réponse. Il est souhaitable de créer des commissions de définition d'objectifs où enseignants, psychologues et spécialistes de la recherche en éducation unissent leurs efforts.

Voici maintenant une série d'*items* de sciences illustrant les principaux niveaux de la hiérarchie de B. Bloom.

TAXONOMIE DE BLOOM

Illustration : Sciences naturelles¹

1. Connaissance.

(a) Connaissance de données spécifiques

Les glucides et les lipides sont utilisés par l'organisme :

- A. Comme aliments énergétiques
- B. Pour former du nouveau cytoplasme
- C. Comme source de sels minéraux
- D. Comme aliments plastiques
- E. Comme source de vitamines

¹ Adapté d'après D.G. LEWIS, Objectives in the Teaching of Science, in *Educational Research*, June 1965, pp. 186-199.

(b) Connaissance de la matière spécifiques

Équilibrez l'équation : H₂ et comptez le nombre

- A. 4
- B. 18
- C. 5
- D. 9
- E. autre nombre que les

(c) Connaissance des principes

Une machine simple ne

- A. Changer la direction
- B. Substituer une grande
- C. Offrir un gain de 100
- D. Vaincre une résistance appliquée
- E. Augmenter l'énergie

2. Compréhension (l'élève comment la matière qui lui a été comment établir des rapports a Si, par temps très froid, poteau en acier, il se peut, verrait pas si le poteau était

C'est parce que l'acier :

- A. A un point de fusion
- B. Rayonne plus la cha
- C. A une masse spéci
- D. Est meilleur conduc
- E. Possède des proprié pas.

3. Application (abstraire d'application l'abstraction d'une au

Quand le magnésium brûle le magnésium, MgO, deux grammes de magnésium dans un récipient : gène :

ant dans le tableau ci-dessus
s d'un cours destiné à l'année
conculaire. Parmi bien d'autres
as n'en avons retenu que trois.

l'ensemble révèle que, même
iées, le professeur se trouve
uels points va-t-il faire porter
tillonnage suffisant de la ma-
e va-t-il réserver aux matières
péciaux qui y correspondent ?
évaluation dans un cours comme

loin d'être simple. Il dépasse
d'un seul homme et débouche
réponse. Il est souhaitable de
ion d'objectifs où enseignants,
e la recherche en éducation

d'items de sciences illustrant
archie de B. Bloom.

DE BLOOM

ces naturelles

spécifiques

sont utilisées par l'organisme :

gétiques

au cytoplasme

s minéraux

ques

aines

s in the Teaching of Science, in Educa-

(b) Connaissance de la manière de traiter des informations
spécifiques

Équilibrez l'équation : $H_2S + O_2 \rightarrow H_2O + SO_2$
et comptez le nombre de molécules pour l'ensemble.

- A. 4
- B. 18
- C. 5
- D. 9
- E. autre nombre que les précédents

(c) Connaissance des principes généraux

Une machine simple ne peut :

- A. Changer la direction d'une force appliquée
- B. Substituer une grande force à une petite force
- C. Offrir un gain de 100 entre les forces mises en jeu
- D. Vaincre une résistance plus grande que la force appliquée
- E. Augmenter l'énergie du système

2. *Compréhension* (l'élève connaît et sait faire usage de la
matière qui lui a été communiquée sans pouvoir nécessaire-
ment établir des rapports avec d'autres matières).

Si, par temps très froid, un enfant porte la main sur un
poteau en acier, il se peut que la main adhère, ce qui n'arri-
verait pas si le poteau était en bois.

C'est parce que l'acier :

- A. A un point de fusion plus élevé que le bois
- B. Rayonne plus la chaleur que le bois
- C. A une masse spécifique plus élevée que le bois
- D. Est meilleur conducteur de la chaleur que le bois
- E. Possède des propriétés magnétiques alors que le bois pas.

3. *Application* (abstraire d'une situation particulière et appli-
quer l'abstraction d'une autre façon).

Quand le magnésium brûle pour former de l'oxyde de ma-
gnésium, MgO , deux grammes d'oxygène se combinent à
3 grammes de magnésium. Si l'on brûle 6 grammes de
magnésium dans un récipient fermé qui contient 3 g d'oxy-
gène :

- A. Il n'y aura pas de réaction
- B. Il se formera du Mg_2O
- C. Il se formera du MgO_2
- D. Une certaine quantité de magnésium restera
- E. L'oxyde de magnésium contiendra un moindre pourcentage d'oxygène

4. *Analyse.*

Si deux objets, X et Y, ont le même moment (momentum), l'énergie cinétique de Y ne peut être supérieure à celle de X que si :

- A. Y est plus lourd que X
- B. Y se déplace plus vite que X
- C. Y a le même poids que X
- D. Y se déplace plus lentement que X
- E. Y est situé à une distance plus grande de l'axe que X

5. *Synthèse* (combiner les éléments pour former un tout structuré).¹

Un certain nombre de problèmes qui ont retenu l'attention des hommes de science peuvent être classés sous la rubrique « structure de la matière ». En ce domaine, les premières recherches portèrent sur la structure discontinue de la matière, hypothèse émise par Dalton en 1805.

Un autre groupe de recherches concerne la théorie cinétique des gaz. Elle apparaît déjà au XVII^e siècle, puis est reprise de façon plus complète par Maxwell (1859) et Boltzmann (1896).

Montrez comment la théorie cinétique des gaz constitua un argument favorable à l'hypothèse de la représentation atomistique de la matière.

Votre discussion doit comprendre :

1. Un exposé des hypothèses fondamentales de cette théorie ;
2. Quelques déductions importantes de cette théorie, appliquée au cas des gaz parfaits ;
3. L'interprétation des lois expérimentales de Gay-Lussac et de l'hypothèse d'Avogadro.

¹ Cet item est dû à H. Guillaume.

6. *Evaluation* (jugements sur à une démonstration ou sur

La théorie atomique de Da

- A. Un bon exemple de
- B. Un bon exemple science
- C. Exacte et a permis
- D. Partiellement fausses vertes scientifiques
- E. Ni correcte, ni inoccupé.

Voici enfin quelques exemples à trois aspects caractéristiques

on

e magnésium restera
contiendra un moindre pour-

e même moment (momentum),
eut être supérieure à celle de

ue X
K
nent que X
ce plus grande de l'axe que X

nts pour former un tout struc-

nes qui ont retenu l'attention
ent être classés sous la ru-
e ». En ce domaine, les pre-
ur la structure discontinue de
ar Dalton en 1805.

es concerne la théorie ciné-
à au XVII^e siècle, puis est re-
par Maxwell (1859) et Boltz-

nétiq ue des gaz constitua un
se de la représentation ato-

re :

ondamentales de cette théo-

tantes de cette théorie, ap-
faits ;

oérimentales de Gay-Lussac
o.

6. *Evaluation* (jugements sur la valeur des arguments servant à une démonstration ou sur des critères).

La théorie atomique de Dalton est :

- A. Un bon exemple de vérité absolue
- B. Un bon exemple d'une conception statique de la science
- C. Exacte et a permis d'autres découvertes scientifiques
- D. Partiellement fausse, mais a permis d'autres découvertes scientifiques
- E. Ni correcte, ni incorrecte, et on ne s'en est plus occupé.

Voici enfin quelques exemples de questions se rapportant à trois aspects caractéristiques du modèle de J.-P. Guilford :

Branche	Production convergente	Production divergente	Evaluation
Sciences	Expliquez pourquoi il ne peut pas y avoir de vie sur Mercure.	En quoi la vie sur Mars pourrait-elle différer de la nôtre ?	Pensez-vous qu'il existe une vie sur Mars ?
Géographie	En quoi le détroit de Bering a-t-il influencé le peuplement de l'Amérique du Nord ?	Que serait-il arrivé si le détroit de Bering n'avait pas existé ?	Actuellement, quel est, à votre avis, le principal rôle du détroit de Bering ?
Histoire	Expliquez l'importance de la découverte de l'Amérique par C. Colomb sur la vie économique de l'Europe.	Que serait-il arrivé si C. Colomb avait découvert la route des Indes au lieu de l'Amérique ?	Quelles sont, à votre avis, les deux conséquences les plus importantes du voyage de C. Colomb ?
Langue maternelle	Expliquez pourquoi la nouvelle s'est beaucoup plus développée aux Etats-Unis qu'en Europe ?	Voici le début d'une nouvelle. Imaginez autant de dénouements que vous le pouvez.	Qu'est-ce qui importe le plus dans une nouvelle : les caractères ou l'intrigue ?

1 D'après J.-R. VERDUIN Jr., *Conceptual Models in Teacher Education*, Washington, A.A.C.T.C., 1967, p. 93.

CHAPITRE REDACTION DES

I. Observations

Pour être équitable et valable, un examen doit toujours comporter un grand domaine à couvrir est étendu, la charge de travail sur une fraction, par exemple, sur une exigence difficilement réalisable dans le système d'enseignement actuel, les élèves sont censés avoir tout appris par eux-mêmes. L'apprentissage est supposée honnêtement arbitrairement sur une partie de l'ensemble pour porter un jugement valable.

En réalité, si, par malchance, une des rares parties qu'un étudiant maîtrise de l'ensemble, par exemple, la grammaire, expliquent une mauvaise réponse à une question bien connue.

Nous avons déjà vu que, parfois, les candidats essaient, en général, de poser une question, sous des formes différentes, pour éviter l'usage dans les examens traditionnels.

Par ailleurs, la multiplicité des questions, ainsi que le modèle de Guilford, sont certainement concevables un petit nombre.

Pourtant, il serait irréaliste d'exiger de l'élève une réponse savante pour le moindre exercice.

Savoir poser des questions est une compétence plus nécessaire au professeur que de corriger les erreurs dans le choix du niveau.

1 Voir 5^e partie : Le mythe de la courbe de

<p>Quelles sont, à votre avis, les deux conséquences les plus importantes du voyage de C. Colomb ?</p>	<p>Qu'est-ce qui importe le plus dans une nouvelle : les caractères ou l'intrigue ?</p>
<p>Colomb avait découvert la route des Indes au lieu de l'Amérique ?</p>	<p>Voici le début d'une nouvelle. Imaginez autant de dénouements que vous le pouvez.</p>
<p>Expliquez pourquoi la nouvelle s'est beaucoup plus développée aux Etats-Unis qu'en Europe ?</p>	<p>Langue maternelle</p>

1 D'après J.-R. VERDUIN Jr., *Conceptual Models in Teacher Education*, Washington, A.A.C.T.C., 1967, p. 93.

CHAPITRE 2.

REDACTION DES QUESTIONS.

I. Observations générales.

Pour être équitable et valide, un examen doit presque toujours comporter un grand nombre de questions. Si le domaine à couvrir est étendu, la chose semble évidente. N'interroger que sur une fraction, parfois minime, de la matière repose sur une exigence difficilement justifiable, surtout dans notre système d'enseignement actuel, non individualisé¹ : tous les élèves sont censés avoir tout appris et la qualité de tous leurs apprentissages est supposée homogène. Des questions portant arbitrairement sur une partie du tout permettraient donc de porter un jugement valable.

En réalité, si, par malchance, le professeur interroge sur une des rares parties qu'un étudiant n'a pas étudiées, un vingtième de l'ensemble, par exemple, c'est la catastrophe. Fréquemment aussi, une simple distraction, une fatigue passagère expliquent une mauvaise réponse dans un domaine pourtant bien connu.

Nous avons déjà vu que, pour prévenir ce danger, les testeurs essaient, en général, de poser au moins trois fois la même question, sous des formes différentes, ce qui n'est certes guère l'usage dans les examens traditionnels.

Par ailleurs, la multiplicité des objectifs à poursuivre, ainsi que le modèle de Guilford nous l'a révélé, rend difficilement concevable un petit nombre de questions.

Pourtant, il serait irréaliste de préconiser une construction savante pour le moindre exercice de contrôle ou de diagnostic.

Savoir poser des questions est probablement la capacité la plus nécessaire au professeur. Mais c'est un art difficile, les erreurs dans le choix du niveau de langage et les obscurités

¹ Voir 5^e partie : Le mythe de la courbe de Gauss.

fréquemment rencontrées dans le libellé des exercices d'application et les questions d'examen en témoignent.

A. Des questions compréhensibles.

Bien des réponses erronées ne sont pas dues à l'ignorance de la matière, mais bien à des malentendus, à une mauvaise compréhension des questions. Le danger apparaît spécialement en un domaine comme les mathématiques parce que trois difficultés s'y combinent aisément : la difficulté inhérente au problème, la difficulté du langage abstrait des mathématiques, et les faiblesses linguistiques des mathématiciens eux-mêmes.

Mais, on le sait, l'abstrus n'épargne pas non plus les historiens, les géographes, les professeurs de sciences... ou de langues.

En ne s'assurant pas de la clarté des questions, les maîtres risquent, en particulier, de commettre une injustice sociale. La recherche contemporaine confirme que beaucoup d'enfants issus de milieux socialement défavorisés souffrent de handicaps graves dans le domaine du langage. Davis et Haggard ont, par exemple, montré qu'il suffit de modifier la forme d'un problème, sans en changer le sens, pour que la différence de réussite entre enfants provenant de milieux socio-économiques favorisés ou non passe de 12 à 32 %¹.

B. Tenir compte du niveau d'information.

Pour vérifier la capacité à transférer les apprentissages, il est souvent nécessaire de situer les problèmes dans un contexte non encore évoqué en classe. Encore faut-il que ce contexte ait un sens pour l'élève.

Tels petits problèmes, récemment rencontrés dans un manuel, supposaient la connaissance des règles du jeu de tennis ou de la manière de remplir des bulletins de pronostic de football. Que signifieront ici les échecs si les élèves n'ont pas la possibilité de s'informer du sens réel des questions avant d'essayer d'y répondre ?

C. Essayer ou prétester les questions.

Autant que possible, il faut essayer les questions avant de les utiliser pour un examen. On découvre ainsi les imprécisions, les défauts de rédaction, les erreurs matérielles et, aussi, le niveau de difficulté.

¹ A. DAVIS, Education for the Conservation of Human Resources, in *Progressive Education*, 1950, 27, 221-224.

Pareil essai est difficile isolé ; de nouveau, le travail est facilité en ce domaine, à l'avance.

D. Calcul de la facilité de

S'il est possible de prévoir le calcul se fera avant l'examen de procéder à l'opération, l'absence de réponses, soit p

Le pourcentage de réussite Rappelons, toutefois, qu'il faut calculer les pourcentages à partir de la somme du nombre de sujets devrait excéder

Une fois l'indice de facilité calculé, comment doser l'examen.

E. Calcul de l'efficacité -

L'indice de facilité seul ne suffit pas, parfois de facteurs accidentels peuvent fausser l'indice d'efficacité. Pour une question déterminée de difficulté faible.

1. Méthode simple.

Pidgeon et Yates proposent une méthode fort utile pour les enseignants.

- Diviser la classe en trois groupes égaux, sur la base des scores.
- Pour chaque *item*, voir quel groupe a réussi.
- Pour chaque *item*, la réussite du groupe supérieur ne donne pas une bonne estimation de la difficulté. Si la différence est faible, le groupe ne distingue pas bien les é

Plus tard, nous verrons comment calculer plus la distribution des résultats (s'aplatit).

libellé des exercices d'appli-
en témoignent.

s.

ne sont pas dues à l'ignorance
malentendus, à une mauvaise
danger apparaît spécialement
matiques parce que trois diffi-
la difficulté inhérente au pro-
strait des mathématiques, et
mathématiciens eux-mêmes.

argne pas non plus les histo-
sseurs de sciences... ou de

rté des questions, les maîtres
être une injustice sociale. La
e que beaucoup d'enfants is-
risés souffrent de handicaps
e. Davis et Haggard ont, par
odifier la forme d'un problè-
que la différence de réussite
socio-économiques favorisés

rmation.

nsférer les apprentissages, il
; problèmes dans un contexte
core faut-il que ce contexte

amment rencontrés dans un
ce des règles du jeu de ten-
es bulletins de pronostic de
hecs si les élèves n'ont pas
ns réel des questions avant

ons.

ssayer les questions avant de
ouvre ainsi les imprécisions,
urs matérielles et, aussi, le

of Human Resources, in *Progressive*

Pareil essai est difficilement réalisable par un enseignant isolé ; de nouveau, le travail en groupe offre bien des possibilités en ce domaine, l'examen étant préparé longtemps à l'avance.

D. Calcul de la facilité des questions.

S'il est possible de prétester sur un grand nombre de sujets, ce calcul se fera avant l'examen. Sinon, il est néanmoins utile de procéder à l'opération, soit pour mieux percevoir la physiologie des réponses, soit pour un usage ultérieur.

Le *pourcentage de réussite* est l'indice le plus simple. Rappelons, toutefois, qu'il n'est pas correct de calculer des pourcentages à partir de petits nombres. Normalement, le nombre de sujets devrait excéder cent.

Une fois l'indice de facilité connu, le professeur sait mieux comment doser l'examen.

E. Calcul de l'efficacité - Pouvoir discriminatif.

L'indice de facilité seul peut induire en erreur, car il résulte parfois de facteurs accidentels. Il est donc prudent de l'accompagner de l'*indice d'efficacité*, qui révèle dans quelle mesure une question déterminée *discrimine* les élèves forts des élèves faibles.

1. Méthode simple.

Pidgeon et Yates proposent une méthode de calcul simple, fort utile pour les enseignants :

- Diviser la classe en trois groupes : supérieur, moyen et inférieur, sur la base des scores totaux à l'examen considéré.
 - Pour chaque *item*, voir quel pourcentage de chacun des trois groupes a réussi.
 - Pour chaque *item*, la différence entre le pourcentage de réussite du groupe supérieur et celui du groupe inférieur donne une bonne estimation de l'efficacité.
- Si la différence est faible ou nulle, c'est que la question ne distingue pas bien les élèves forts des élèves faibles.

Plus tard, nous verrons que, plus la discrimination est fine, plus la distribution des résultats est large (la courbe de Gauss s'aplatit).

2. Méthode plus fine.

Voici une autre méthode, utilisée pour l'examen de fin du secondaire en Angleterre¹.

Quatre groupes A, B, C, D sont constitués.

On donne les consignes suivantes :

- 1) Déterminer, pour chacun des *items* et dans chacun des quatre groupes, le nombre d'élèves qui répondent correctement.
- 2) Calculer la moyenne obtenue par chacun des quatre groupes pour chaque *item* ou question.
- 3) Réunir toutes ces données en un tableau d'ensemble et comparer les moyennes pour chaque *item* dans les quatre groupes.

Si les quatre moyennes s'ordonnent de la même façon que les scores moyens pour le test entier, dans les quatre groupes, on peut affirmer que l'*item* considéré contribue à la discrimination totale du test. Sinon, l'*item* est suspect.

Exemple :

Question n°	1	2	3	4	5	6	7	Total
Maximum	6	6	6	6	10	8	8	50
Groupe A	4,9	5,1	5,3	5,2	4,3	5,8	4,3	34,9
Groupe B	3,2	4,9	5,8	4,1	4,7	5,9	3,8	32,3
Groupe C	3,4	4,6	5,4	3	3,4	3,6	3,9	27,3
Groupe D	2	3,4	5	2,8	2,3	4,1	2,3	21,8
Ordre des scores moyens	A	A	B	A	B	B	A	A
	C	B	C	B	A	A	C	B
	B	C	A	C	C	D	B	C
	D	D	D	D	D	C	D	D
Discrimine	Non	Oui	Non	Oui	Non	Non	Non	Oui

La réponse que l'on obtient ainsi manque de finesse. Pour les *items* considérés comme acceptables, nous ne pouvons pas dire s'ils discriminent finement ou si, au contraire, ils sont tout

¹ D. MATHER et al., *The C.S.E., A Handbook for Moderators*, Londres, Collins, 1965, p. 108.

² A partir d'un ensemble de 80 copies, D. Mather constitue quatre groupes de 20 : A = les 20 premiers ; B = les 20 suivants ; etc.

juste acceptables. De plus, nous ne pouvons pas améliorer ou la rédaction de la question.

Pour obtenir des informations plus précises, nous utilisons des techniques compliquées. La méthode suivante est simple et apporte les résultats.

Considérons les résultats obtenus par les élèves de chaque groupe. Nous les présentons dans un tableau pour les questions 2, 3 et 4. Au lieu d'indiquer les scores, nous les pointons dans un tableau.

Question 2 — Maximum 6				
Scor. obt.	Groupes			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Question 4	
Scor. obt.	A
	6
5	
4	
3	
2	
1	
0	

Scor. obt : scores obtenus.

¹ Encore que l'utilisation de l'ordinateur.

isée pour l'examen de fin du

nt constitués.

ntes :

items et dans chacun des
èves qui répondent correcte-

par chacun des quatre grou-
estion.

n un tableau d'ensemble et
chaque item dans les quatre

donnent de la même façon
e test entier, dans les quatre
l'item considéré contribue à
t. Sinon, l'item est suspect.

4	5	6	7	Total
6	10	8	8	50
5,2	4,3	5,8	4,3	34,9
4,1	4,7	5,9	3,8	32,3
3	3,4	3,6	3,9	27,3
2,8	2,3	4,1	2,3	21,8
A	B	B	A	A
B	A	A	C	B
C	C	D	B	C
D	D	C	D	D
Oui	Non	Non	Non	Oui

si manque de finesse. Pour
ables, nous ne pouvons pas
si, au contraire, ils sont tout

for Moderators, Londres, Collins, 1965,

lather constitue quatre groupes de 20 :

juste acceptables. De plus, nous ne savons trop comment
améliorer ou la rédaction de la question, ou la façon de l'évaluer.

Pour obtenir des informations plus fines, les statisticiens
utilisent des techniques compliquées¹. La solution graphique
suivante est simple et apporte les nuances souhaitées.

Considérons les résultats individuels de chacun des 20
élèves de chaque groupe. Nous prendrons pour exemple les
questions 2, 3 et 4. Au lieu d'écrire les résultats en chiffres,
nous les pointons dans un tableau.

Question 2 — Maximum 6				
Scor. obt.	Groupes			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Question 3 — Maximum 6				
Scor. obt.	Groupes			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Question 4 — Maximum 6				
Scor. obt.	Groupes			
	A	B	C	D
6				
5				
4				
3				
2				
1				
0				

Scor. obt : scores obtenus.

¹ Encore que l'utilisation de l'ordinateur les mette aujourd'hui à la portée de chacun.

es, nous indiquons le quartile inférieur¹, le comportement

Question 3				
Points	A	B	C	D
6	QS	QS	QS	QS
5	M	M	M	M
4	QI	QI	QI	QI
3				
2				
1				
0				

4		
B	C	D
S		QS
M	QS	
	M	
QI		M
	QI	
		QI

s mesures, nous devrions
a A en D. La descente est
pouvoir discriminatif. Une

le de notes ordonnées ; si le nombre
étique entre les deux notes médianes

le point au-dessus duquel se trouvent
u milieu de la moitié supérieure de la
milieu de la moitié inférieure.

montée indique qu'un groupe inférieur (d'après le résultat à l'ensemble de l'examen) obtient de meilleurs points que le groupe supérieur.

On observe :

Question 2 :

— Il n'y a aucune remontée et une tendance à la descente, mais pas très nette.

Item faiblement discriminatif.

Question 3 :

— Aucune descente pour le quartile supérieur.

— Remontée pour le médian.

— Remontée pour le quartile inférieur.

Item non discriminatif. A rejeter.

Question 4 :

— Aucune remontée.

— Descente bien marquée pour le médian.

— Descente très bien marquée pour le quartile inférieur.

Bonne discrimination. Une réserve, cependant, pour le QS du groupe D : les correcteurs ont été trop généreux pour les meilleurs élèves du groupe faible (D).

II. Réponses ouvertes ou fermées ?

A une question à *réponse ouverte*, l'élève répond spontanément, en utilisant son propre vocabulaire. La *réponse* est dite *fermée* si le sujet est tenu d'opérer un choix parmi plusieurs réponses proposées.

Traditionnellement, l'école s'en est exclusivement tenue au premier type de questions ; elle négligeait ainsi un outil de grande utilité.

A. Réponses ouvertes (orales ou écrites).

Ce sont les questions les plus naturelles, celles que nous posons à tout instant dans la vie.

Elles conviennent spécialement, soit pour des épreuves de contrôle faites rapidement, en toute spontanéité par les maîtres, en cours d'enseignement, soit pour la vérification d'apprentissages tellement complexes qu'ils échappent à l'analyse rigoureuse.

L'évaluation des capacités supérieures (créativité, jugement, esprit critique,...) semble de leur ressort. Pas exclusivement, toutefois. D'abord parce que, comme nous allons le voir dans un instant, les questions à choix multiple bien construites permettent des explorations beaucoup plus subtiles qu'il n'y paraît. Ensuite, parce que, par souci d'objectivité, de rigueur, les notateurs tendent peut-être, de façon inconsciente, à ne retenir des réponses ouvertes que les éléments les plus concrets, les plus factuels qu'elles contiennent. La divergence des évaluations n'est-elle pas fonction de la subtilité de l'objet sur lequel elles portent ?

« Nous devons admettre, » écrit Vernon¹, « que les élèves excellents ont l'occasion de montrer certaines qualités exceptionnelles dans l'examen traditionnel, et qu'un notateur perspicace peut, parfois, s'en apercevoir, alors que les autres peuvent très bien pénaliser la réponse à cause de son anticonformisme. » Le débat est loin d'être clos.

En construisant entièrement la réponse, l'élève tente de prouver deux choses : sa connaissance de la matière et sa capacité à l'exprimer verbalement.

Longtemps, on a cru que les deux allaient de pair. Le vieil adage : « Ce que l'on conçoit bien s'énonce clairement, Et les mots pour le dire arrivent aisément », fut vérité reçue par bien des maîtres jusqu'aujourd'hui. Pourtant, les choses sont loin d'être aussi simples. La traduction verbale de la pensée n'est qu'une forme d'expression parmi d'autres, dédaignées pendant des siècles par les classes sociales privilégiées auxquelles tout travail manuel, tout traitement du réel répugnaient.

Renoncer entièrement aux questions à réponses ouvertes serait une erreur dans une civilisation où la communication verbale reste dominante. Mais il est parfois bon d'isoler, au moins partiellement, les connaissances et les capacités d'expression verbale, ce que permettent les réponses fermées. Les professeurs doivent d'autant moins hésiter à y recourir que, comme le remarquent Pidgeon et Yates, on n'a jamais démontré que leur utilisation, même fréquente, nuisait au développement verbal.

¹ Bull. 4, p. 7.

Une réponse formulée en toute spontanéité est souvent inconvenue et par conséquent souvent inconvénient grave : son caractère subjectif est évident. L'évaluation par comparaison avec les autres est souvent délicate. C'est assurément une des principales raisons pour lesquelles, entre correcteurs, si souvent désaccordés, on trouve souvent de grandes divergences.

B. Réponses fermées. Questions

1. Utilité.

L'examen intensif vise à vérifier les acquisitions ; il a souvent un but précis et porte sur une matière vaste.

Dans les deux cas, de nombreux sujets sont proposés. D'où le recours aux questions à choix multiple, surtout du type « à choix multiple ».

Exemple : Le premier test de la pratique courante a été :

- a) Binet
- b) Galton
- c) Goddard
- d) Spearman
- e) Terman.

2. Constituer une provision de questions

Trouver plusieurs dizaines de questions pour un seul examen, met l'imagination à l'épreuve. Un fichier d'items que l'on enrichit régulièrement, permet de considérer la tâche.

On n'inscrit qu'une question à choix multiple, mais on porte aussi les « distracteurs » possibles, mais vraisemblables, à côté de la question. Les distracteurs sont notamment choisis fréquemment par les élèves ; ils sont choisis pour leur « logique » et la vraisemblance.

Les professeurs d'une matière doivent unir leurs efforts en cette matière.

On semble aussi s'acheminer vers la constitution de fichiers d'items, offices centraux pour les maîtres plusieurs centaines de questions mises au point pour des populations.

soit pour des épreuves de spontanéité par les maîtres, la vérification d'apprentissage se rapporte à l'analyse rigou-

érieures (créativité, jugement, leur ressort. Pas exclusivement, comme nous allons le voir, à choix multiple bien que beaucoup plus subtiles par souci d'objectivité, de rigueur, de façon inconsciente, que les éléments les plus importants contiennent. La divergence n'est pas de la subtilité de l'objet

Vernon¹, « que les élèves possèdent certaines qualités exceptionnelles et qu'un notateur perspicace ne peut pas les reconnaître sans se servir de son anticonformisme »

réponse, l'élève tente de donner une réponse de la matière et sa

exemples allaient de pair. Le vieil homme énonçait clairement, Et les choses furent véritablement reçues par bien des esprits, les choses sont loin d'être banales de la pensée n'est pas dédaignée pendant les siècles privilégiées auxquelles elle est si réellement répugnante.

On passe à des questions à réponses ouvertes où la communication est parfois bon d'isoler, au lieu de les capacités d'exécution des réponses fermées. Les professeurs ont tendance à y recourir que, comme on n'a jamais démontré que cela favorisait le développement

Une réponse formulée en toute liberté présente un autre inconvénient grave : son caractère unique se prête mal à l'évaluation par comparaison avec les réponses d'autres individus. C'est assurément une des principales sources des désaccords entre correcteurs, si souvent dénoncés par la docimologie.

B. Réponses fermées. Questions (*items*) à choix multiple.

1. *Utilité.*

L'examen intensif vise à vérifier en détail la qualité des acquisitions ; il a souvent un but diagnostique. L'examen extensif porte sur une matière vaste.

Dans les deux cas, de nombreuses vérifications sont nécessaires. D'où le recours aux questions à « réponses fermées », surtout du type « à choix multiple ».

Exemple : Le premier test d'intelligence utilisable dans la pratique courante a été construit par :

- a) Binet
- b) Galton
- c) Goddard
- d) Spearman
- e) Terman.

2. *Constituer une provision de questions.*

Trouver plusieurs dizaines de questions précises, pour un seul examen, met l'imagination à rude épreuve. Constituer un fichier d'*items* que l'on enrichit à chaque occasion facilite considérablement la tâche.

On n'inscrit qu'une question par fiche afin de pouvoir porter aussi les « distracteurs », c'est-à-dire les réponses fausses, mais vraisemblables, à mesure de leur découverte. Ces distracteurs sont notamment fournis par les erreurs commises fréquemment par les élèves ; on est ainsi assuré de respecter leur « logique » et la vraisemblance est garantie.

Les professeurs d'une même branche peuvent aisément unir leurs efforts en cette matière.

On semble aussi s'acheminer vers la création de *banques d'items*, offices centraux pouvant mettre à la disposition des maîtres plusieurs centaines de questions à choix multiple, bien mises au point pour des populations déterminées.

3. Exploiter la gamme des possibilités logiques.

Répondre à un *item* à choix multiple n'est pas nécessairement un simple exercice de mémoire. loin s'en faut.¹

Les recherches entreprises par les services de l'enseignement supérieur en vue de la réforme des examens de médecine, en France, ont bien mis en lumière la richesse du système². Nous choisissons à dessein des exemples dans le domaine des examens universitaires, dans l'espoir de convaincre les maîtres que, quelle que soit la matière qu'ils enseignent, le recours à l'examen objectif n'est pas exclu *a priori*.

Dans le présent examen, on distingue huit types de questions :

a) Question à complément simple.

L'*item* se présente, dans son esprit sinon toujours dans sa forme, comme une phrase à compléter :

Ex. : Parmi les caractères suivants, celui qui s'applique à tous les enzymes est :

- A. Ils contiennent toujours un coenzyme dissociable.
- B. Ils sont thermostables.
- C. Ils contiennent toujours de l'azote dans leur molécule.
- D. Ils contiennent toujours du phosphore dans leur molécule.
- E. Ils sont dialysables.

Remarquons qu'il n'est pas toujours nécessaire que la réponse correcte figure parmi les choix proposés. C aurait pu prendre la place de E et devenir : « Aucune des réponses précédentes. »

L'*item* peut prendre la forme négative :

Ex. : Un hydrosol métallique a tous les caractères suivants, sauf un. Indiquez lequel.

A — B — C — D — E

Pour répondre, le candidat doit connaître tous les caractères de l'hydrosol. Toutefois, la forme négative contraint à une gymnastique intellectuelle qui se superpose à la difficulté inhérente à la matière.

¹ La technique a même été appliquée avec succès à l'analyse littéraire - voir B. CHOPPIN et A. PURVES, *A comparison of open-ended and multiple choice items dealing with literary understanding in Research in the Teaching of English*, 3, 1, 1969, 15-24.

² Paris, Ministère de l'Education nationale. Enseignement supérieur, Examens et concours, section médicale 1961. Tous les exemples médicaux suivants sont empruntés à cette publication.

b) Association simple.

Elle sert à vérifier la correspondance d'entités qui peuvent être classées.

Il faut ici faire correspondre une lettre à un élément précédent (voir l'exemple ci-dessous).

Ex. : Branche postérieure

- A. Segment d'origine
- B. Groupe des rameaux
- C. Groupe des rameaux
- D. Nerf interosseux postérieur
- 5. est appliqué sur la face antérieure de l'os osseux
- 6. contourne le col du radius
- 7. innerve les muscles de la région antibrachiale
- 8. passe entre les deux os
- 9. innerve les muscles de la région antibrachiale postérieure

c) Association composée

C'est une simple variante de l'association simple.

Exemple :

- A. Paludisme à plasmodium falciparum
- B. Paludisme à plasmodium vivax
- C. Les deux à la fois (A et B)
- D. Aucun des deux

Question 1 :

- 1. Le développement du paludisme est moins grave chez un homme de race blanche que chez un homme de race noire.
- 2. Une association de médicaments est recommandée pour le traitement de choix.
- 3. Les épisodes cliniques du paludisme traité de chloroquine une fois par semaine.
- 4. Guérit d'une manière définitive le paludisme traité de chloroquine (réponse C).
- 5. La contamination est plus grave chez un homme traité de chloroquine une fois par semaine que chez un homme traité de chloroquine deux fois par semaine.

Remarquons que le nombre de contaminations est augmenté ou diminué selon l'intensité du traitement.

ilités logiques.

multiple n'est pas nécessaire-
aire, loin s'en faut. ¹

ar les services de l'enseigne-
e des examens de médecine,
re la richesse du système ².
exemples dans le domaine
l'espoir de convaincre les
matière qu'ils enseignent, le
pas exclu *a priori*.

istingue huit types de ques-

emple.

esprit sinon toujours dans sa
ter :

vants, celui qui s'applique à

n coenzyme dissociable.

l'azote dans leur molécule.
phosphore dans leur molé-

jours nécessaire que la ré-
hoix proposés. C aurait pu
Aucune des réponses précé-

négative :

tous les caractères suivants,

- D — E

it connaître tous les carac-
me négative contraint à une
superpose à la difficulté inhé-

succès à l'analyse littéraire - voir B.
nded and multiple choice items dealing
aching of English, 3, 1, 1969, 15-24.
Enseignement supérieur, Examens et
es médicaux suivants sont empruntés à

b) Association simple.

Elle sert à vérifier la connaissance « d'un certain nombre
d'entités qui peuvent être ou ne pas être en relation ».

Il faut ici faire correspondre un élément précédé d'une
lettre à un élément précédé d'un chiffre (association ou appa-
riement).

Ex. : *Branche postérieure du nerf radial*

- A. Segment d'origine
- B. Groupe des rameaux postérieurs
- C. Groupe des rameaux antérieurs
- D. Nerf interosseux postérieur
5. est appliqué sur la face postérieure du ligament inter-
osseux
6. contourne le col du radius
7. innerve les muscles de la couche superficielle de la
région antibrachiale postérieure
8. passe entre les deux chefs du court supinateur
9. innerve les muscles de la couche profonde de la région
antibrachiale postérieure

c) Association composée.

C'est une simple variante de la précédente.

Exemple :

- A. Paludisme à plasmodium vivax
- B. Paludisme à plasmodium falciparum
- C. Les deux à la fois (A et B)
- D. Aucun des deux

Question 1 :

1. Le développement clinique a toutes chances d'être
moins grave chez un homme de race noire que chez un
homme de race blanche (réponse B)
2. Une association de primaquine et de chloroquine est le
traitement de choix pour une attaque aiguë (réponse A)
3. Les épisodes cliniques sont supprimés par l'ingestion
de chloroquine une fois par semaine en zone endémique
(réponse C)
4. Guérit d'une manière définitive par le traitement avec la
chloroquine (réponse B)
5. La contamination est évitée par l'ingestion de chloro-
quine une fois par semaine (réponse D)

Remarquons que le nombre de questions peut être ici aug-
menté ou diminué selon l'importance accordée à la matière.

phénomènes numérotés sont
Il faut indiquer le trouble
pond pas (2).

diagnostique
diagnostique
diagnostique

se à effet.

ale supérieure permet des
és PARCE QUE la tête du
ment annulaire.

n proposée sont toutes les
e relation de cause à effet

n proposée sont toutes les
de relation de cause à effet

nais la raison proposée est

mais la raison proposée est
é.

n proposée sont toutes les

met le candidat dans une
réelle.

de est un homme de 21 ans
ne toux et de fièvre.

s avant l'admission par un
ctoration, suivis dans les 24
nt de 37.8 à 38.3 qui a per-
mission.

ie, la toux s'accroît, pro-
xpectoration blanche et vis-

Trois jours avant l'admission, des accès paroxystiques de toux commencèrent, parfois suivis de vomissements. Des sensations de frissons furent notées, mais non pas de véritables frissons avec tremblements. Une douleur parasternale antérieure à la toux existe depuis le cinquième jour de maladie.

A l'examen physique, la température est à 38.3, le pouls à 110, le rythme respiratoire 32, la tension maxima 10 1/2, minima 8.

Le malade est bien développé, sans maigreur, sa maladie semble aiguë, il est dyspnéique, mais non cyanosé.

L'examen physique de la cage thoracique montre des vibrations vocales à la palpation et à l'auscultation. Le murmure vésiculaire est normal. Dans l'aisselle gauche, on entend quelques râles fins et la qualité bronchique du son est augmentée, bien que d'intensité normale.

La formule sanguine est la suivante : globules blancs 3 400 (polynucléaires 30 %, lymphocytes 62 %, monocytes 5 %, éosinophiles 3 %).

La radio du thorax révèle une augmentation de la densité de la région périhilaire avec des aires mal définies de densité inégale, nuageuse aux deux bases et dans un champ pulmonaire supérieur gauche.

Questions :

1. *Quel est le diagnostic le plus probable ?*
 - a) tuberculose
 - b) pneumonie à pneumocoques
 - c) pneumonie (primaire atypique) à virus
 - d) coccidiomycose
 - e) broncho-pneumonie
2. *Quel est le signe physique qui s'y ajoute probablement ?*
 - a) spénomégalie
 - b) signe de souffrance méningée
 - c) bruit de frottements pleuraux
 - d) changements fréquents dans la distribution des symptômes thoraciques
 - e) signe de condensation lobaire gauche
3. *Lequel des examens de laboratoire suivants va de pair avec le diagnostic ?*
 - a) l'élévation et l'augmentation des agglutines froides

- b) hémoculture positive
- c) leucocytose marquée au début de la convalescence
- d) examen des expectorations
- e) cuti-réaction positive

4. *Quelle thérapeutique devra être employée ?*

- a) repos au lit et streptomycine
- b) repos au lit et pénicilline
- c) streptomycine et acide para-amino salicylique
- d) repos au lit et auréomycine
- e) psychothérapie et rééducation physique

5. *Quelle est l'issue probable de cette maladie sans traitement?*

- a) la fièvre va disparaître spontanément par une crise terminale
- b) la convalescence va être progressive avec une rechute prévisible
- c) un empyème résiduel va se développer
- d) une fibrose résiduelle va apparaître après guérison
- e) une caverne pulmonaire peut apparaître

g) Comparaisons quantitatives.

Comparez X à Y :

X — pression mécanique dans le capillaire veineux
 Y — pression oncotique dans le capillaire veineux

Et dites si :

- A. X est plus grand que Y
- B. Y est plus grand que X
- C. X est égal à Y

h) Relations.

Soit :

1. Le débit circulaire cutané — ET
 2. La quantité de chaleur perdue par unité de temps
- A. L'augmentation du premier est accompagnée d'une augmentation du second, ou, la diminution du premier est accompagnée d'une diminution du second.
 - B. L'augmentation du premier est accompagnée d'une diminution du second, ou, la diminution du premier est accompagnée d'une augmentation du second.
 - C. Les variations du second sont indépendantes des variations du premier.

i) Compléments groupés

Procédé utilisé quand une réponse correcte.

Ex. : Cinq conscrits mesurent 1,63 m et 1,61 m.

1. La moyenne des tailles
 2. L'écart type est proche
 3. L'écart type est proche
 4. L'échantillon a de bonnes caractéristiques pour la population des tailles
- A. 1, 2 et 3 sont corrects
 - B. 1 et 3 sont corrects
 - C. 2 et 4 sont corrects
 - D. 4 est correct
 - E. Une seule des propositions est correcte

On trouvera, en annexe, un exemple de question conduite selon la méthode de choix multiple.

4. Calcul de l'efficacité des distracteurs

Dans une question à choix multiples, les distracteurs jouent bien leur rôle si, pour chaque question, que l'on propose une possibilité de réponse a reçu

a) Situation idéale.

A	B	C
10%	10%	10%

D est la bonne réponse.
 Les autres choix ont été

b) Situation à corriger.

A	B	C
3%	0%	35%

Seul, le distracteur C a un effet lent sans pouvoir.

ut de la convalescence

employée ?

mino salicylique

physique

te maladie sans traitement?

anément par une crise ter-

gressive avec une rechute

velopper

raître après guérison

apparaître

e capillaire veineux

e capillaire veineux

ET
e par unité de temps
est accompagnée d'une
u, la diminution du premier
minution du second.
est accompagnée d'une di-
diminution du premier est
tation du second.
nt indépendantes des varia-

i) Compléments groupés.

Procédé utilisé quand une question peut avoir plus d'une réponse correcte.

Ex. : Cinq conscrits mesurent : 1,65 m - 1,67 m - 1,69 m - 1,63 m et 1,61 m.

1. La moyenne des tailles de l'échantillon est 1,65 m
 2. L'écart type est proche de 8
 3. L'écart type est proche de 2,8
 4. L'échantillon a de bonnes chances de renseigner sur la population des tailles des individus en général
- A. 1, 2 et 3 sont corrects
 - B. 1 et 3 sont corrects
 - C. 2 et 4 sont corrects
 - D. 4 est correct
 - E. Une seule des propositions 1, 2 ou 3 est correcte

On trouvera, en annexe, une comparaison entre un examen conduit selon la méthode traditionnelle et selon la méthode à choix multiple.

4. Calcul de l'efficacité des distracteurs.

Dans une question à choix multiple, il importe de vérifier si les distracteurs jouent bien leur rôle. A cet effet, on calcule, pour chaque question, quel pourcentage des choix chaque possibilité de réponse a recueilli.

a) Situation idéale.

A	B	C	D	E
10 %	10 %	10 %	60 %	10 %

D est la bonne réponse.

Les autres choix ont été également attractifs.

b) Situation à corriger.

A	B	C	D	E
3 %	0 %	35 %	60 %	2 %

Seul, le distracteur C a réellement joué. Les autres se révèlent sans pouvoir.

5. Critiques et réfutation partielle.

Bien que sa valeur soit prouvée depuis longtemps, l'examen par questions à choix multiple continue à susciter de vives critiques. Certaines peuvent être aisément réfutées ; d'autres paraissent justifiées ; d'autres encore ne peuvent être ni infirmées, ni confirmées, faute de critères scientifiques.

a) Une objectivité trompeuse.

La notation des questions à choix multiple est indiscutablement objective : les réponses correctes sont définies d'avance ; l'élève les a trouvées ou non. Les avantages de cette méthode sont évidents.

Pourtant, la subjectivité est loin d'avoir été complètement éliminée : elle subsiste, au moins en partie, dans la rédaction des questions et dans la décision concernant la réponse à considérer comme correcte.

Dans la rédaction des questions d'abord. Elles sont le fruit de la réflexion, de l'invention des examinateurs qui, en dernière analyse, agissent subjectivement. A côté des questions auxquelles ils ont pensé, combien d'autres, peut-être plus valides, n'aurait-on pas pu imaginer ?

Dans la rédaction des réponses considérées comme idéales, ensuite. On a rarement affaire à des réponses aussi évidemment correctes que $2 \times 2 = 4$. Au cours de la construction de tests avec nos étudiants, nous avons plus d'une fois rencontré des réponses proposées comme correctes, passées victorieusement au crible de l'analyse mathématique, et qui, pourtant, n'étaient, au mieux, que grossières approximations exprimées dans un langage douteux.

Comme Vernon le note, non sans malice, la supériorité indiscutable des examens « objectifs » sur les épreuves traditionnelles est sans doute plus souvent due à la préparation très soignée et à l'union des efforts et de la compétence de nombreux enseignants et psychopédagogues, qu'à la nature même de l'examen¹.

On oublie aussi qu'entre l'utopie de l'objectivité et la subjectivité totale, bien des stades intermédiaires existent.

¹ VERNON, The C.S.E.: *An Introduction to Objective-Type Examinations*, Londres, H.M.S.O., 1964, pp. 4-5.

b) Choix « corrects » co

Ce point aussi a été e
qui écrit :

« Il est fréquent que de
accueillent les questions à
telles que : « Cet *item* est stu
considérés comme « faux » p
sibles, sinon plus, que la rép
vient principalement du fait
sur les mêmes capacités que
demment, des examens con
contenir pas mal de mauvais
épreuves construites par de
ment, car les *items* insatisfai
contrôle préliminaire, par de
cernée, ou bien éliminés lo
réponses (...). Les critiques
réactions sophistiquées peuv
élèves sur lesquels les *item*
ques peuvent lire dans les
nent même pas à l'esprit d'é

c) Un jeu de hasard.

Dans une question ferm
correcte et l'autre fausse, la
qui travaille au pur hasard
constructeurs de tests le sav

Aujourd'hui, on a le plu
cinq choix, dont l'un est corr
blables (distracteurs)². Dan
considérablement réduit. La
leurs selon leur niveau de co
de la loi de la probabilité, l
chance minimale de tomber a
Par contre, une connaissanc
ment un certain nombre de
dans le choix aveugle parm
beaucoup plus grande que
tice.

¹ VERNON, o.c., p. 6.

² Parfois aussi, plusieurs choix co
sélectionner le plus adéquat.

b) Choix « corrects » contestables.

Ce point aussi a été excellemment discuté par Vernon¹ qui écrit :

« Il est fréquent que des personnes hautement éduquées accueillent les questions à choix multiple par des critiques telles que : « Cet *item* est stupide » ou prétendent que les choix, considérés comme « faux » par le testeur, sont tout aussi admissibles, sinon plus, que la réponse dite « bonne ». Cette critique vient principalement du fait que l'*item* objectif ne s'appuie pas sur les mêmes capacités que les questions traditionnelles. Evidemment, des examens construits par des amateurs peuvent contenir pas mal de mauvais *items*. On en trouve aussi dans les épreuves construites par des professionnels, mais plus rarement, car les *items* insatisfaisants sont ou bien rejetés, lors du contrôle préliminaire, par des spécialistes de la discipline concernée, ou bien éliminés lors de l'analyse mathématique des réponses (...). Les critiques risquent aussi d'oublier que leurs réactions sophistiquées peuvent différer très fort de celles des élèves sur lesquels les *items* ont été essayés. Enfin, les critiques peuvent lire dans les questions des choses qui ne viennent même pas à l'esprit d'élèves intelligents. »

c) Un jeu de hasard.

Dans une question fermée à deux choix de réponses, l'une correcte et l'autre fausse, la probabilité de réussite de l'élève qui travaille au pur hasard est, théoriquement, de 50 %. Les constructeurs de tests le savent depuis longtemps.

Aujourd'hui, on a le plus souvent recours aux questions à cinq choix, dont l'un est correct et les autres seulement vraisemblables (distracteurs)². Dans ce cas, le jeu du hasard est considérablement réduit. La démarche des élèves variera d'ailleurs selon leur niveau de connaissance de la matière. En vertu de la loi de la probabilité, l'étudiant qui ignore tout n'a qu'une chance minime de tomber aveuglément sur la solution correcte. Par contre, une connaissance partielle permet d'éliminer sciemment un certain nombre de distracteurs ; la chance de réussite dans le choix aveugle parmi les possibilités restantes est alors beaucoup plus grande que dans le cas précédent, et c'est justice.

¹ VERNON, o.c., p. 6.

² Parfois aussi, plusieurs choix conviennent, mais à des degrés divers. Il faut alors sélectionner le plus adéquat.

On peut d'ailleurs réduire largement le rôle de la chance en corrigeant les résultats obtenus à l'aide d'une formule simple¹, laquelle, il est vrai, pénalise généralement de façon exagérée, puisque la part du choix aveugle varie selon les élèves.

d) Acrobatie mentale.

On reproche aux examens traditionnels de faire la part trop belle à la facilité d'expression. On peut toutefois se demander si les *items* à choix multiple un peu compliqués (nous venons d'en rencontrer des exemples) ne présentent pas un inconvénient au moins aussi grave en ajoutant, à la difficulté inhérente à la matière de la question, l'obligation de démêler des doubles négations, de saisir des subtilités logiques ou, plus généralement, en accordant une prime à une certaine aptitude à l'abstraction à partir de données verbales. P. Vernon constate, en tout cas, que la compréhension de la lecture joue un rôle important dans la réussite de ces épreuves².

Il importe, non seulement de connaître le niveau de développement mental des élèves, mais aussi de savoir clairement ce que l'on veut : vérifier la connaissance de la matière, la capacité en compréhension de la lecture et en raisonnement, ou la combinaison des deux.

On ne voit pas comment le sondage des qualités intellectuelles supérieures irait sans un accroissement de la complexité des *items*. C'est pourquoi plusieurs auteurs estiment que les questions ouvertes continueront à jouer un rôle important dans les examens de niveau avancé.

En cas de recours à la correction automatique, une autre difficulté vient encore s'ajouter : l'utilisation de cartes de réponses où ne figurent généralement que les numéros des questions et les lettres A, B, C, D, E, représentant les cinq choix. Ce système exige un supplément d'attention de la part de l'élève, en particulier quand il ne peut répondre à certaines questions et décide de les sauter. S'il oublie de sauter aussi la ligne correspondante sur la carte de réponses, les conséquences peuvent être désastreuses. La difficulté ne doit toutefois pas être exagérée : Remmers, Gage et Rummel ont montré expérimentalement que le système est utilisable avec des enfants, à partir de 9 ou 10 ans³.

$$1 \text{ Nombre de réponses correctes} - \frac{\text{Nombre de réponses incorrectes}}{\text{Nombre de choix} - 1}$$

2 P. VERNON, The Determinants of Reading Comprehension, in *Educational Psychological Measurement*, 1962, 22, 269-286.

3 H. REMMERS, N. GAGE et RUMMEL, *Educational Measurement and Evaluation*, New York, Harper, 1955, p. 246.

De toute façon, les élèves entraînés à la technique de subir une épreuve décisive, construction rigoureuse, les bons résultats, leur utilisation, depuis de nombreuses années, répandue ailleurs, le prouve.

e) Inconvénients incertains

Nous qualifions les réponses que, à notre connaissance, aucune n'en a établi le bien-fondé ne se justifient que par ce qui est construit.

- Choisir la bonne réponse et construire. La mémoire intellectuelle, une répétition claire de la pensée.
- Sachant que, pour subir, il suffirait de souligner ou de mettre une croix, les maîtres ne s'opposent pas à la solution.
- La préférence irait à de nouvelles questions en un temps très court ; on exige une très longue durée et la vie réelle ne nous permet pas.

C. En guise de conclusion

Un compromis nous semble le moins mauvais.

Puisque les questions ouvertes possèdent chacune des avantages, cause certains apprentissages, raison de revendiquer un rôle, peuvent parfaitement coexister, estiment cependant qu'il faut tester les deux types de questions.

Par ailleurs, une forme mixte d'examens ouverts et fermés existe, plus, selon les progrès de la recherche.

L'examen consisterait en questions ouvertes, d'une portée limitée, et questions fermées testées. L'éventail des réponses.

ment le rôle de la chance à l'aide d'une formule généralement de façon variable selon les élèves.

onnels de faire la part trop toutefoix se demander si compliqués (nous venons présentent pas un inconvénient, à la difficulté inhérente de démêler des doubles logiques ou, plus généralement une certaine aptitude à l'abstraction. P. Vernon constate, en la lecture joue un rôle important².

connaître le niveau de développement aussi de savoir clairement l'importance de la matière, la capacité et en raisonnement, ou la

développement des qualités intellectuelles et de la complexité des auteurs estiment que les auteurs jouent un rôle important dans

on automatique, une autre utilisation de cartes de mémoire que les numéros des cartes représentant les cinq choix. L'attention de la part de l'élève peut répondre à certaines questions s'il oublie de sauter aussi de réponses, les conséquences de la difficulté ne doit toute-fois et Rummel ont montré qu'il est utilisable avec des en-

réponses incorrectes

re de choix — 1
compréhension, in *Educational Psychological Measurement and Evaluation*,

De toute façon, les élèves doivent être soigneusement entraînés à la technique des examens à choix multiple avant de subir une épreuve décisive. Moyennant cette précaution, et une construction rigoureuse, les questions à choix multiple donnent de bons résultats, leur utilisation intensive dans les pays anglo-saxons, depuis de nombreuses années, et de plus en plus répandue ailleurs, le prouve à suffisance.

e) Inconvénients incertains.

Nous qualifions les reproches suivants d'incertains parce que, à notre connaissance, aucune recherche scientifique rigoureuse n'en a établi le bien-fondé. Il semble que certaines critiques ne se justifient que dans la mesure où l'examen est mal construit.

- Choisir la bonne réponse parmi d'autres est plus facile que la construire. La mémoire intervient trop. Une certaine paresse intellectuelle, une répugnance à l'effort nécessaire, à la formulation claire de la pensée risquent de s'installer.
- Sachant que, pour subir l'examen en langue maternelle, il suffirait de souligner ou d'indiquer la réponse choisie par une croix, les maîtres négligeraient les exercices d'expression.
- La préférence irait à de nombreux petits exercices à faire en un temps très court ; on négligerait ainsi les problèmes qui exigent une très longue réflexion. Or, les études supérieures et la vie réelle ne nous épargnent pas ce genre d'épreuves.

C. En guise de conclusion : un compromis.

Un compromis nous paraît s'imposer à deux niveaux au moins.

Puisque questions ouvertes et questions fermées semblent posséder chacune des avantages particuliers et mettre en cause certains apprentissages différents, on ne voit aucune raison de revendiquer un monopole pour l'une des deux. Elles peuvent parfaitement coexister. Les meilleurs docimologistes estiment cependant qu'il n'est pas souhaitable de mélanger les deux types de questions dans une même épreuve.

Par ailleurs, une formule intermédiaire entre questions ouvertes et fermées existe et s'imposera peut-être de plus en plus, selon les progrès de la recherche pédagogique.

L'examen consisterait en un nombre assez élevé de questions *ouvertes*, d'une portée relativement limitée, ayant été pré-testées. L'éventail des réponses probables serait donc connu

d'avance, ce qui permettrait de proposer un schéma de notation s'approchant très fort de la rigueur de la question à réponse fermée.

Cette possibilité est déjà confirmée par plusieurs travaux expérimentaux. Ainsi, des corrélations presque parfaites (.98 et .99) ont pu être obtenues entre sept groupes de notateurs représentant chacun une commission d'examens de Grande-Bretagne. L'expérience a été faite sur des épreuves d'histoire de niveau supérieur du G.G.E. (fin du secondaire) ¹.

Sans aller aussi loin, des professeurs peuvent parfaitement s'entendre sur des points à exiger dans les réponses à des questions ouvertes.

Par ailleurs, la façon de rédiger les questions peut fermer partiellement des réponses et donc diminuer le jeu de la subjectivité. Une question comme : Ecrivez un début approprié à la phrase : « ... quand il commença à pleuvoir. » est, en quelque sorte, à mi-chemin entre l'épreuve subjective et l'épreuve objective. Car ici, les cas de désaccord entre notateurs invités à se prononcer sur la correction de la réponse seront rares.

¹ JOINT MATRICULATION BOARD, *The Marking of Scripts in Advanced Level History*. Universities of Manchester, Liverpool, Leeds, Sheffield and Birmingham, 1964.

I. Un préambule indispo

Il est impossible de dis
sans s'appuyer sur quelq
n'oublions pas la promesse
l'arithmétique élémentaire no

A. La courbe de Gauss, im

Un sac contient deux
sinon que l'une est rouge et

Dans ces conditions, ch
de sortir à chaque tirage ave

Si nous nous livrons à c
il est possible que nous tir
même couleur. A mesure qu
tirages, cette probabilité dim

Imaginons 100 tirages c
le rouge sortira à peu près
par contre, fort peu probabl
vement la boule rouge ou la
petit peu plus probable que
fois blanc, plus probable d
etc. Bref, la probabilité va
blanc, puis diminuer progres
99 blanc.

oser un schéma de notation
r de la question à réponse

rmée par plusieurs travaux
ns presque parfaites (.98 et
groupes de notateurs repré-
amens de Grande-Bretagne.
reuves d'histoire de niveau
re) ¹.

seurs peuvent parfaitement
dans les réponses à des

r les questions peut fermer
diminuer le jeu de la sub-
vez un début approprié à la
pleuvoir. » est, en quelque
bjective et l'épreuve objec-
entre notateurs invités à se
ponse seront rares.

CHAPITRE 3.

LA NOTATION.

I. Un préambule indispensable : la courbe de Gauss.

Il est impossible de discuter de l'évaluation des résultats sans s'appuyer sur quelques notions mathématiques. Nous n'oublions pas la promesse faite au début de cet ouvrage : l'arithmétique élémentaire nous suffira !

A. La courbe de Gauss, image de la probabilité.

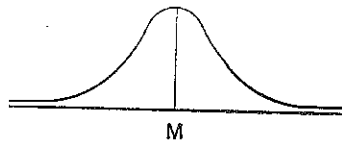
Un sac contient deux boules, en tout point semblables, sinon que l'une est rouge et l'autre blanche.

Dans ces conditions, chaque boule a une chance sur deux de sortir à chaque tirage aveugle.

Si nous nous livrons à ce jeu, un très petit nombre de fois, il est possible que nous tirions plusieurs fois consécutives la même couleur. A mesure que nous augmentons le nombre de tirages, cette probabilité diminue.

Imaginons 100 tirages consécutifs. Il est fort probable que le rouge sortira à peu près autant de fois que le blanc. Il est, par contre, fort peu probable que l'on tirera 100 fois consécutivement la boule rouge ou la boule blanche. Il est déjà un tout petit peu plus probable que l'on obtiendra 99 fois rouge et une fois blanc, plus probable déjà 98 fois rouge et 2 fois blanc, etc. Bref, la probabilité va augmenter jusqu'à 50 rouge et 50 blanc, puis diminuer progressivement pour arriver à 1 rouge et 99 blanc.

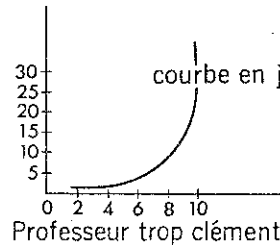
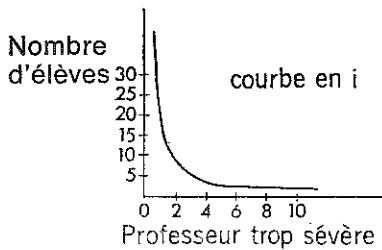
Dans un très grand nombre de tirages, ce mouvement ascendant-descendant correspond à une courbe revêtant la forme d'une cloche : c'est la fameuse courbe de Gauss, dont les deux moitiés sont symétriques par rapport à la moyenne arithmétique et dont les extrémités ne touchent jamais la ligne du zéro, la probabilité nulle n'existant qu'à l'infini.



Cette *distribution*, dite *normale*, est à l'image de beaucoup de qualités humaines, telles qu'elles se répartissent dans des groupes nombreux, *pris au hasard*. Ainsi, parmi les habitants d'une grande ville, les hommes de taille moyenne sont les plus nombreux, tandis que les géants et les nains sont très rares ; entre ces deux extrêmes, la population se distribue selon la courbe de Gauss.

Même l'erreur se soumet souvent à la loi normale : si l'on fait un très grand nombre de mesures, on verra fort probablement apparaître une erreur de grandeur moyenne, et une distribution allant de l'erreur infime à l'erreur maximum. Nous verrons, par la suite, que cette observation est importante.

Dès que le hasard ne joue plus, la distribution se modifie. Un professeur peut, par exemple, donner une très grande majorité de mauvaises notes et fort peu de bonnes, ou le contraire. Il est possible que l'on obtienne alors une des deux courbes caractéristiques suivantes :



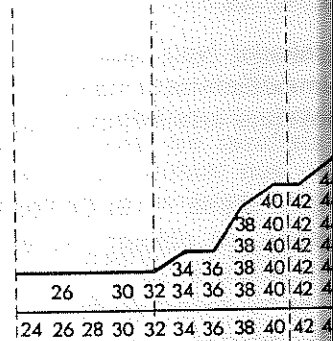
B. La courbe de Gauss, I de l'enseignement non

Un professeur qui en dans une classe où les sélectionnés, donne norm difficulté adapté à la maj correct, il y aura donc be très bons et peu de très tats s'approchera de la co

Pareil phénomène se dans les exercices échap parce qu'ils mettent en je

Ci-dessous, on a disp nus à un examen, lui-même ves ont obtenu 50 sur 10 obtient la note la plus ba

On observe aussi qu tion des résultats n'est pa seulement son ébauche. donner quelques précisi le nombre de sujets ava polie et on considère la

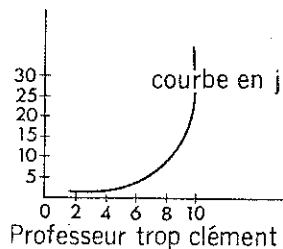


de tirages, ce mouvement
à une courbe revêtant la
e courbe de Gauss, dont les
rapport à la moyenne arith-
ouchent jamais la ligne du
u'à l'infini.

est à l'image de beaucoup
s se répartissent dans des
Ainsi, parmi les habitants
e taille moyenne sont les
nts et les nains sont très
population se distribue selon

nt à la loi normale : si l'on
es, on verra fort probable-
eur moyenne, et une distri-
rreur maximum. Nous ver-
ion est importante.

la distribution se modifie.
donner une très grande
rt peu de bonnes, ou le
tienne alors une des deux



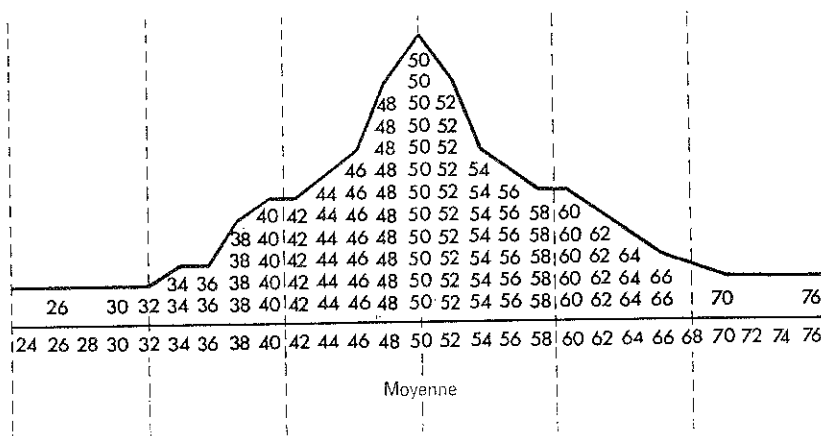
B. La courbe de Gauss, image des résultats de l'enseignement non individualisé.

Un professeur qui enseigne de façon non individualisée dans une classe où les élèves ne sont pas spécialement sélectionnés, donne normalement à son cours un degré de difficulté adapté à la majorité du groupe. Si l'ajustement est correct, il y aura donc beaucoup de résultats moyens, peu de très bons et peu de très mauvais. La distribution de ces résultats s'approchera de la courbe gaussienne.

Pareil phénomène se produit, plus spontanément encore, dans les exercices échappant à la quantification rigoureuse, parce qu'ils mettent en jeu un ensemble complexe de facteurs.

Ci-dessous, on a disposé 100 résultats (imaginaires) obtenus à un examen, lui-même noté sur 100. On constate que 12 élèves ont obtenu 50 sur 100 (note moyenne), alors qu'un seul obtient la note la plus basse (26) et un seul la plus haute (76).

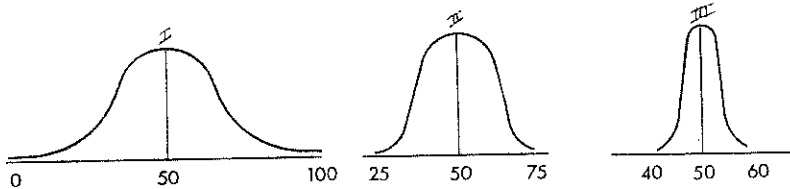
On observe aussi que la ligne correspondant à la répartition des résultats n'est pas une courbe en cloche parfaite, mais seulement son ébauche. Dans pareil cas (nous allons bientôt donner quelques précisions à ce propos), on suppose que, si le nombre de sujets avait été plus élevé, la courbe se serait polie et on considère la répartition comme « normale ».



C. L'écart type ou sigma, indice précieux.

1. Signification.

La variation des 100 résultats que nous venons d'examiner aurait pu être soit plus grande, par exemple de 0 à 100, soit plus petite, par exemple de 40 à 60, tout en se distribuant toujours en courbe de Gauss. Nous aurions pu avoir :



Dans les trois cas, la moyenne est 50. Pourtant, il s'agit de situations très différentes.

La marge de variation des résultats est :

$$\text{courbe I : } 100 - 0 = 100$$

$$\text{courbe II : } 75 - 25 = 50$$

$$\text{courbe III : } 60 - 40 = 20$$

Sigma ou σ , est un indice facile à calculer (voir plus loin la méthode de calcul), qui nous indique immédiatement comment les résultats varient par rapport à la moyenne arithmétique des notes obtenues.

Pour les 100 notes prises en exemple, $\sigma = 9$. Si l'on prend un σ avant et après la moyenne, on obtient : 41 et 59. Comptez le nombre de résultats qui tombent entre ces deux limites, vous obtenez 68 notes, soit ici 68 % de notes de tout le groupe.

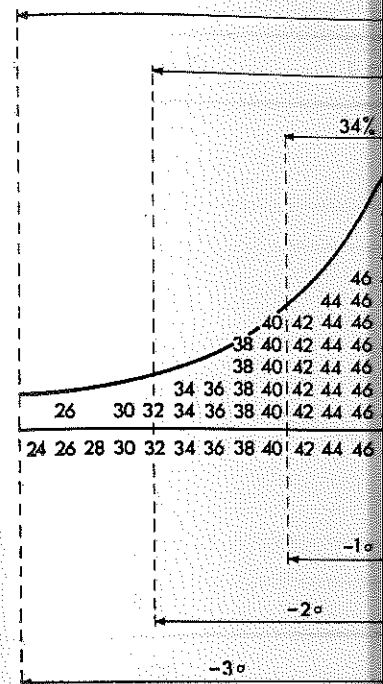
Il en est toujours ainsi, dans une *distribution normale* : le σ indique toujours la même proportion des résultats par rapport à l'ensemble :

$$1 \text{ sigma de part et d'autre de la moyenne} = 68 \% \text{ des notes}$$

$$2 \text{ sigma de part et d'autre de la moyenne} = 95 \% \text{ des notes}$$

$$3 \text{ sigma de part et d'autre de la moyenne} = 99 \% \text{ des notes.}$$

Voyons maintenant la distribution de quelques nouveaux :



Comme, à un σ donné, la surface de l'aire déterminée est la même proportion des résultats, on peut comparer les résultats obtenus, pour autant que ces résultats soient normalement distribués. Nous montrerons bientôt comment.

2. Estimation rapide de la moyenne

a) Problème.

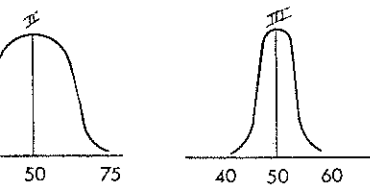
Le calcul rigoureux de la moyenne (méthode de Gauss) est relativement lourd. Dans une distribution normale, la moyenne arithmétique est égale à la tendance centrale.

Un procédé simple permet de fournir que des résultats approximatifs dans presque tous les cas.

1. D'après D.R. MATHER et al., o.c.

précieux.

que nous venons d'examiner par exemple de 0 à 100, soit 60, tout en se distribuant toujours nous pu avoir :



le est 50. Pourtant, il s'agit de

ultats est :

- 0 = 100
- 25 = 50
- 40 = 20

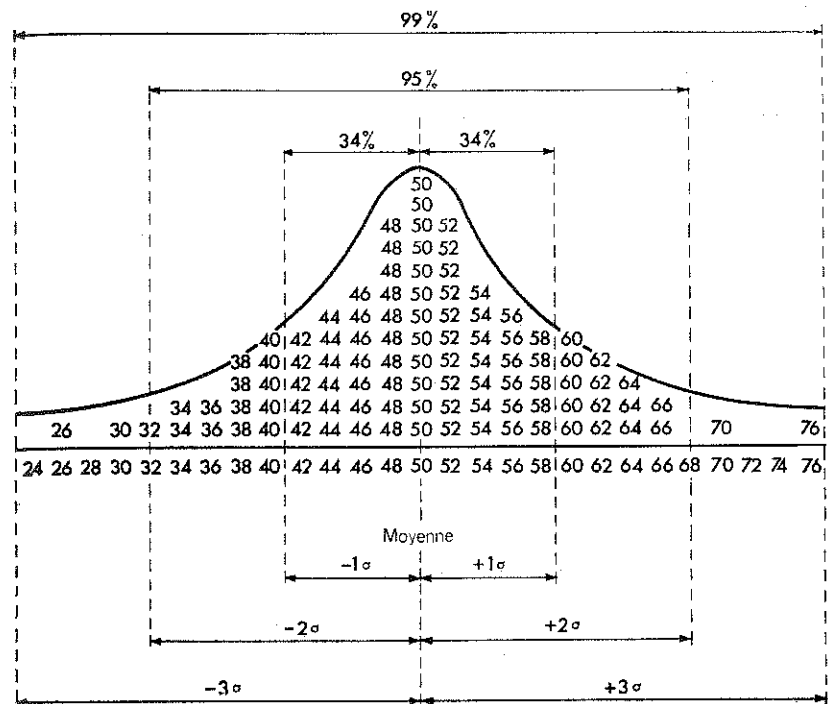
ile à calculer (voir plus loin la que immédiatement comment la moyenne arithmétique des

exemple, $\sigma = 9$. Si l'on a moyenne, on obtient : 41 et ts qui tombent entre ces deux it ici 68 % de notes de tout

une *distribution normale* : le proportion des résultats par

- moyenne = 68 % des notes
- moyenne = 95 % des notes
- moyenne = 99 % des notes.

Voyons maintenant la distribution des résultats avec des yeux nouveaux :



Comme, à un sigma donné, correspond toujours la même surface de l'aire déterminée par la courbe, c'est-à-dire la même proportion des résultats, *cet indice nous fournit le moyen de comparer les résultats attribués par des professeurs différents, pour autant que ces résultats se répartissent normalement*. Nous montrerons bientôt comment.

2. Estimation rapide de la moyenne et du sigma.

a) Problème.

Le calcul rigoureux de la moyenne et de l'écart type (sigma) est relativement lourd. De plus, quand la distribution n'est pas normale, la moyenne arithmétique donne une idée fautive de la tendance centrale.

Un procédé simple permet d'éviter de longs calculs. Il ne fournit que des résultats *approximatifs*. Ils suffisent néanmoins dans presque tous les cas de la pratique courante à l'école¹.

¹ D'après D.R. MATHER et al., o.c.

Soit deux questions d'examen auxquelles 10 élèves ont répondu. Voici les notes attribuées, les moyennes et les écarts types calculés par la méthode classique.

Elève	Question 1	Question 2
1	39	32
2	33	28
3	25	32
4	22	28
5	26	27
6	18	31
7	23	33
8	13	27
9	57	35
10	45	36
	301	309

Moyenne arithmétique : 30

Ecart type :

$$\sigma = \sqrt{\frac{\sum d^2}{N}} = \pm 13.$$

Moyenne arithmétique : 31

Ecart type :

$$\sigma = \pm 3.$$

b) Méthode simple de calcul

1° Représentation graphique

Notes	Question 1	Question 2
60		
58		
56		
54		
52		
50		
48		
46		
44		
42		
40		
38		
36		
34		
32		
30		
28		
26		
24		
22		
20		
18		
16		
14		
12		
10		

Remarquons que, dans ce cas, la moyenne et le médian sont égaux.

2° Estimation de l'écart type

$$\sigma = \frac{3}{4} \text{ de l'écart interquartile}$$

Donc :

a. Chercher le quartile supérieur des notes au-dessus du médian

$$\text{Question 1} = 39$$

$$\text{Question 2} = 33$$

amen auxquelles 10 élèves ont
uées, les moyennes et les écarts
classique.

Question 1	Question 2
39	32
33	28
25	32
22	28
26	27
8	31
3	33
3	27
7	35
5	36

1 309
Moyenne arithmétique : 31
Ecart type : $\sigma = \pm 3$.

b) *Méthode simple de calcul.*

1° Représentation graphique des résultats.

Notes	Question 1	Question 2
60		
58	.	
56		
54		
52		
50		
48		
46	.	
44		
42		
40		
38	.	:
36		:
34	.	:
32		:
30		
28		: :
26	:	
24		
22	:	
20		
18	.	
16		
14	.	
12		
10		

Marges de variation :

Question 1 : $57 - 13 = 44$
Question 2 : $36 - 27 = 9$

Les notes de la question 1 sont beaucoup plus dispersées. Elles influencent plus le classement final que celles de la question 2. Il est souvent utile d'ignorer les deux notes extrêmes : dans ce cas :

Marge Question 1 = 27
Question 2 = 8

Médians

= notes du milieu.

Ici, chiffre pair, donc moyenne entre 5^e et 6^e.

Question 1 :

$(25 + 26) : 2 = 25,5$

Question 2 :

$(31 + 32) : 2 = 31,5$

Pour la question 1, la différence assez nette entre la moyenne arithmétique et le médian (30 et 25,5) indique une distribution asymétrique des notes (ici : majorité sous le médian).

Pour la question 2, la symétrie est bonne (Moy. 31 et Méd. 31,5).

Remarquons que, dans une distribution parfaitement normale, la moyenne et le médian coïncident : voir notre exemple.

2° Estimation de l'écart type (σ).

$\sigma = 3/4$ de l'écart interquartile.

Donc :

a. Chercher le quartile supérieur, c'est-à-dire le milieu des notes au-dessus du médian.

Question 1 = 39

Question 2 = 33

b. Chercher le quartile inférieur.

$$\text{Question 1} = 22$$

$$\text{Question 2} = 28$$

c. Ecart interquartile.

$$\text{Question 1} = 39 - 22 = 17$$

$$\text{Question 2} = 33 - 28 = 5$$

d. Estimation du σ :

$$\text{Question 1} = 17 \times \frac{3}{4} = 12 \frac{3}{4}$$

$$\text{Question 2} = 5 \times \frac{3}{4} = 3 \frac{3}{4}$$

D. La concentration des résultats autour de la moyenne.

Reportons-nous de nouveau à la distribution des 100 notes prises comme exemple (p. 97). Nous observons que, plus nous nous rapprochons de la moyenne, plus les notes sont nombreuses :

12 notes sur 100 à la moyenne exacte

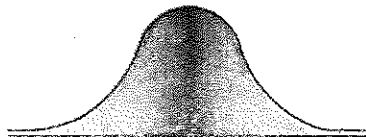
22 notes sur 100 entre 48 et 52

68 notes sur 100 entre 41 (-1σ) et 59 ($+1\sigma$)

Imaginons qu'il s'agisse de compositions françaises. Alors que le professeur pouvait noter de 0 à 100, presque un quart des notes se situent entre 48 et 52, soit à 4 points de distance.

Que signifie une différence de 4 points sur 100 en dissertation française ? Il est à peu près certain que, si nous faisons juger le même travail par dix professeurs différents, les écarts seront presque toujours supérieurs à 2. Or, si c'est la moyenne qui constitue la limite de l'échec (ici, elle coïncide avec la moitié de l'échelle totale), on peut dire, sans crainte de se tromper que, pour 22% du groupe d'élèves pris comme exemple, l'échec ou la réussite dépendra du pur hasard.

Voici, en dégradé, une image de la densité de population par rapport à la moyenne :



E. Courbe de Gauss voulue par

La courbe de Gauss est un élément important puisqu'elle permet d'identifier (entre 25 et 35) et les moins bons.

Dans le cas d'un concours postulent un petit nombre de personnes, il faut choisir parmi ceux qui se situent à l'extrême qui échappe à la forte concentration de notes. même, si l'on souhaite classer les candidats trop exclusivement dans le passé, une courbe normale est utile pour permettre à chacun d'être évalué. normale est certainement utile.

Comme il ne s'agit plus, ce n'est pas du hasard, mais de sujets spécialement créés artificiellement pour créer des conditions gaussiennes. Comment ? En dosant la difficulté (et d'efficacité).

F. Comment savoir si une distribution

Tout ce que nous venons de dire sur la courbe de Gauss et du sigma n'est valable que si la distribution est normale. Avant d'engager une telle démarche, il faut donc vérifier si cette condition est remplie.

Il existe des procédés mathématiques pour atteindre ce but, mais, de nouveau, on se contente d'une simple approximation graphique.

rieur.

= 17

= 5

= $12 \frac{3}{4}$

= $3 \frac{3}{4}$

ts autour de la moyenne.

à la distribution des 100 notes
ous observons que, plus nous
ne, plus les notes sont nom-

e exacte

2
1 σ) et 59 (+ 1 σ)

compositions françaises. Alors
de 0 à 100, presque un quart
52, soit à 4 points de distance.

le 4 points sur 100 en disserta-
s certain que, si nous faisons
fesseurs différents, les écarts
rs à 2. Or, si c'est la moyenne
ec (ici, elle coïncide avec la
eut dire, sans crainte de se
groupe d'élèves pris comme
épendra du pur hasard.

e de la densité de population



E. Courbe de Gauss voulue par les maîtres.

La courbe de Gauss est un excellent instrument de classe-
ment puisqu'elle permet d'identifier les meilleurs (par exemple
entre 25 et 35) et les moins bons.

Dans le cas d'un concours où de nombreux candidats
postulent un petit nombre de places, on peut, par exemple,
choisir parmi ceux qui se situent entre + 1 σ et + 3 σ (partie
qui échappe à la forte concentration autour de la moyenne). De
même, si l'on souhaite classer les élèves — ce que l'on a fait
trop exclusivement dans le passé, mais qui reste tout de même
utile pour permettre à chacun de se situer —, la distribution
normale est certainement utile.

Comme il ne s'agit plus, cette fois, de groupes choisis au
hasard, mais de sujets spécialement entraînés, le professeur
crée artificiellement les conditions nécessaires à une répartition
gaussienne. Comment ? En dosant des *items* selon leurs indices
de difficulté (et d'efficacité).

F. Comment savoir si une distribution est normale ?

Tout ce que nous venons de dire à propos de la courbe
de Gauss et du sigma n'est valable que si les résultats se distri-
buent *normalement*. Avant d'engager toute opération, il importe
donc de vérifier si cette condition est remplie.

Il existe des procédés mathématiques rigoureux pour
atteindre ce but, mais, de nouveau, nous nous contenterons ici
d'une simple approximation graphique : l'*histogramme*.

Le procédé est simple :

1. *Classer les résultats.*

On peut les ordonner du plus grand au plus petit, ou inversement.

Voici ce qu'on obtient en classant les 100 résultats que nous avons déjà pris comme exemples :

Notes	Nombre d'élèves ayant obtenu ces notes (fréquences) (1)
26	1
30	1
32	1
34	2
36	2
38	4
40	5
42	5
44	6
46	7
48	10
50	12
52	10
54	7
56	6
58	5
60	5
62	4
64	3
66	2
70	1
76	1

$N = 100$

Mais, si les notes sont nombreuses, ce procédé n'est ni rapide, ni pratique. Il est bien plus aisé de constituer des classes.

- 1° Calculer la marge de variation entre les deux notes extrêmes : $76 - 26 = 50$.
- 2° Diviser cette marge par 15^1 : $50 : 15 = 3,33$.
- 3° Choisir comme intervalle de classe un des deux nombres impairs les plus proches : 3 ou 5. On choisit ici 5, vu le petit nombre de notes.

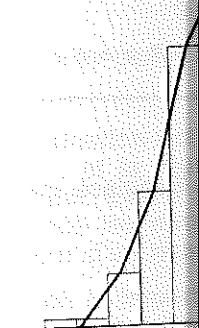
1 Ce nombre est arbitraire. La pratique montre que, dans la plupart des cas, il conduit à une bonne répartition.

4° Placer la note supérieure rieur : note supérieure : donc : 74-75-76-77-78.

	Classes
1	74 - 78
2	69 - 73
3	64 - 68
4	59 - 63
5	54 - 58
6	49 - 53
7	44 - 48
8	39 - 43
9	34 - 38
10	29 - 33
11	24 - 28
11 classes	Intervalle (i) = 5

2. *Dessiner l'histogramme.*

Dans la partie « fréquences » page horizontalement pour graphique appelé histogramme.



grand au plus petit, ou inversement les 100 résultats que nous avons :

Nombre d'élèves ayant obtenu ces notes (fréquences) (f)

1	1
1	1
1	1
2	2
2	2
4	4
5	5
5	5
6	6
7	7
10	10
12	12
10	10
7	7
6	6
5	5
5	5
4	4
3	3
2	2
1	1
1	1

$N = 100$

reuses, ce procédé n'est ni plus aisé de constituer des

on entre les deux notes extrê-

$1 : 50 : 15 = 3,33.$

le classe un des deux nombres : 3 ou 5. On choisit ici 5.

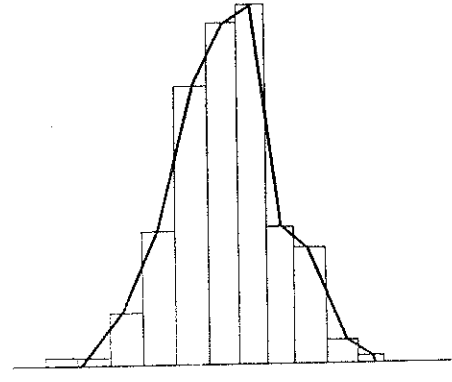
ontre que, dans la plupart des cas, il

4° Placer la note supérieure au milieu de l'intervalle supérieur : note supérieure : 76 ; la classe supérieure est donc : 74-75-76-77-78.

	Classes	Fréquences (f)
1	74 - 78	I 1
2	69 - 73	I 1
3	64 - 68	IIII 5
4	59 - 63	IIII IIII 9
5	54 - 58	IIII IIII IIII III 18
6	49 - 53	IIII IIII IIII IIII II 22
7	44 - 48	IIII IIII IIII IIII III 23
8	39 - 43	IIII IIII 10
9	34 - 38	IIII III 8
10	29 - 33	II 2
11	24 - 28	I 1
11 classes	Intervalle (i) = 5	Nombre de notes (N) = 100

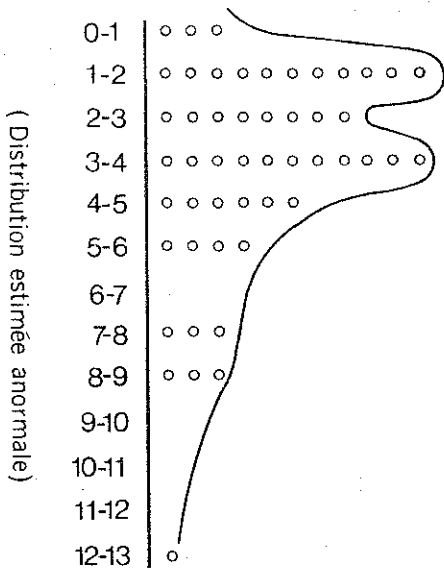
2. Dessiner l'histogramme.

Dans la partie « fréquences » du tableau ci-dessus, le pointage a été fait avec beaucoup de soin. Il suffit de tenir la page horizontalement pour voir apparaître le squelette du graphique appelé histogramme : il reste à tracer les rectangles.

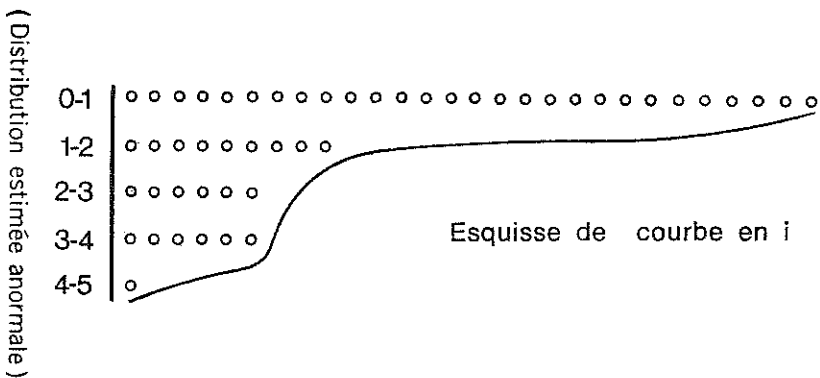


Si l'on joint les milieux des bases supérieures, on obtient une ligne (polygone de fréquences) qui, ici, est une ébauche dont on se contentera pour considérer que la distribution est fort probablement normale. Pourquoi ? Parce que le tracé n'a qu'un *sommet*, à peu près *centré*, à partir duquel les quantités *décroissent assez régulièrement de part et d'autre*.

Par contre, les deux distributions ci-dessous ne seraient pas considérées comme *normales*.



Esquisse de courbe décentrée et à deux sommets (courbe bimodale).



Esquisse de courbe en i

II. La notation subjective

A. Introduction.

Par notation subjective, non faite par les maîtres, d'une façon lors de l'observation de la performance avoir eu une vue d'ensemble sur la façon plus générale, est subjective arrive par une démarche non automatique.

Le degré de subjectivité variable elle joue presque sans limite (C'est très bien... ») ; tantôt, elle cherche objective (exemple : échelle avec une précision ne laissant l'interprétation personnelle).

Dans tous les cas, l'évaluation par comparaison à des critères externes à l'égalité ou à l'inégalité par rapport à un classement.

C'est pourquoi nous pouvons parler des échelles d'évaluation.

L'échelle d'évaluation aide à classer les objets par rapport à une qualité plus ou moins élevée.

Elle représente un continuum d'un minimum absolu à un maximum, n'est presque toujours qualitatif absolu. De l'excellence... à... l'inférieur grand, etc.

L'échelle d'évaluation, grande

soit *numérique* : 1
0
soit *descriptive* : Très bien E

L'échelle d'évaluation est employé à l'école... et le plus crainte de se tromper, qu'il a l'habitude des examens, à tous les niveaux ment fait le malheur d'un nom

C'est pourquoi il importe de connaître sa véritable nature et surtout les

bases supérieures, on obtient (encens) qui, ici, est une ébauche considérer que la distribution ale. Pourquoi ? Parce que le eu près centré, à partir duquel sez régulièrement de part et

utions ci-dessous ne seraient ales.

Esquisse de courbe décentrée et à deux sommets (courbe bimodale).

Esquisse de courbe en i

II. La notation subjective : l'échelle d'évaluation.

A. Introduction.

Par notation subjective, nous entendons toute évaluation faite par les maîtres, d'une façon globale, soit immédiatement lors de l'observation de la performance des élèves, soit après avoir eu une vue d'ensemble sur des scores particuliers. De façon plus générale, est subjective toute note à laquelle on arrive par une démarche non automatisable.

Le degré de subjectivité varie considérablement. Tantôt, elle joue presque sans limite (par exemple : « C'est bien — C'est très bien... ») ; tantôt, elle cède presque le pas à la démarche objective (exemple : échelle dont les degrés sont définis avec une précision ne laissant pratiquement plus de place à l'interprétation personnelle).

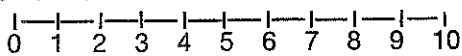
Dans tous les cas, l'évaluation subjective se fait par comparaison à des critères externes ou internes. Le notateur conclut à l'égalité ou à l'inégalité par rapport au critère. Il opère donc un classement.

C'est pourquoi nous pouvons limiter cette partie à l'étude des échelles d'évaluation.

L'échelle d'évaluation aide à ordonner des individus ou des objets par rapport à une qualité qu'ils possèdent à un degré plus ou moins élevé.

Elle représente un *continuum* qui peut théoriquement aller d'un minimum absolu à un maximum absolu, mais qui, pratiquement, n'est presque toujours qu'une fraction, un segment de cet absolu. De l'excellence... à... l'insuffisance, du très petit... au très grand, etc.

L'échelle d'évaluation, graphique ou non, est :

soit *numérique* : 
soit *descriptive* : Très bien Bien Moyen Faible Insuffisant



L'échelle d'évaluation est l'instrument le plus fréquemment employé à l'école... et le plus mal connu. On peut dire, sans crainte de se tromper, qu'il a littéralement vicié une large partie des examens, à tous les niveaux de la scolarité, et probablement fait le malheur d'un nombre considérable d'individus.

C'est pourquoi il importe absolument d'en connaître la véritable nature et surtout les immenses limitations.

B. Nature et faiblesse des échelles d'évaluation.

L'échelle d'évaluation est une échelle ordinale.

- 1) Elle n'a ni zéro ni maximum naturels ou absolus ; elle commence et finit n'importe où, au choix de l'utilisateur. Par conséquent, même s'ils évaluent un même phénomène, du même point de vue, les juges n'ont jamais des échelles rigoureusement les mêmes ; les différences peuvent être considérables, et le sont souvent.

Exemple : *Connaissance de l'anglais deuxième langue.*

Théoriquement, le zéro pourrait être le moment où l'élève ne connaît pas encore un seul mot de la langue. Toutefois, on juge rarement à partir de ce point à l'école.

Quant à la connaissance totale, absolue, elle n'existe évidemment pas.

Imaginons deux professeurs d'anglais jugeant, sans se concerter, deux classes où chaque élève a fait un court exposé oral. Fort probablement, les deux extrémités des échelles d'évaluation utilisées indépendamment par les deux professeurs seront plus définies par l'élève jugé le plus fort et l'élève jugé le plus faible, dans son groupe, que par une évaluation abstraite de la quantité d'anglais qui devrait être connue au niveau pédagogique considéré.

- 2) Les degrés ne sont pas de même grandeur à l'intérieur d'une même échelle.

Il est, par exemple, impossible de démontrer que la distance séparant la bonne de la très bonne connaissance de l'anglais est la même que la distance séparant la connaissance moyenne de la bonne.

Si les degrés sont : A, B, C, D, E, on ne peut donc pas dire que $A - B = B - C = 0$

- 3) Des degrés correspondants d'une échelle à l'autre (pour une même branche) ne sont pas de même grandeur.

Bon en anglais, pour un professeur, n'est pas quantitativement égal à bon en anglais pour un autre : rien ne peut établir pareille égalité de façon mathématique.

- 4) Des échelles portant sur des pas comparables.

Pour les élèves d'une même classe, le seul maître enseigne toutes les matières. La notation en mathématique ne recouvre pas le bon en français. Les échelles ne sont pas comparables sur le plan quantitatif.

De tout ceci, il résulte que le système ne met en évidence aucune opération arithmétique. La notation française et l'anglaise ne permettent pas d'additionner un savoir total. C'est pourquoi, après dix ans d'enseignement, on ne peut pas dire que l'élève a acquis un savoir total. C'est pourquoi, dans nos écoles, on ne peut pas, plus, d'utiliser des échelles différentes branches, en français, en mathématique : de 0 à 120 pour les sciences, etc.

Si le système a pu fonctionner, c'est parce qu'il a permis d'identifier les meilleurs (et non pas tout dans les premiers) et les moins bons.

C. Utilité.

Malgré ses limitations, l'échelle est, dans des cas, le seul instrument qui permet de mesurer notre jugement sur des phénomènes complexes.

En fait, elle permet d'apprécier des éléments qualitatifs, aussi bien que des éléments quantitatifs. Elle est un complément des résultats des examens objectifs ou les tests.

D. Construction.

Beaucoup d'éducateurs se sont efforcés de construire une échelle apparente avec laquelle on compare les résultats.

En réalité, il n'est pas rare que l'instrument exige de longs mois de clarification théorique, et que les méthodes raffinées comme l'analyse factorielle ne permettent pas de déterminer dans quelle mesure les mêmes choses sous des

es d'évaluation.

échelle ordinale.

aturels ou absolus ; elle com-
au choix de l'utilisateur. Par
nt un même phénomène, du
s n'ont jamais des échelles
es différences peuvent être
t.

anglais deuxième langue.

ait être le moment où l'élève
eul mot de la langue. Toute-
artir de ce point à l'école.
ale, absolue, elle n'existe évi-

d'anglais jugeant, sans se
haque élève a fait un court
nt, les deux extrémités des
ndépendamment par les deux
s par l'élève jugé le plus fort
ans son groupe, que par une
tité d'anglais qui devrait être
considéré.

ne grandeur à l'intérieur d'une

de démontrer que la distance
bonne connaissance de l'an-
nce séparant la connaissance

E, on ne peut donc pas dire

ne échelle à l'autre (pour une e même grandeur.

asseur, n'est pas quantitative-
pour un autre : rien ne peut
n mathématique.

4) Des échelles portant sur des branches différentes ne sont pas comparables.

Pour les élèves d'une même classe d'école primaire où un seul maître enseigne toutes les branches, être *bon* en mathématique ne recouvre pas les mêmes phénomènes qu'être *bon* en français. Les échelons ne sont certainement pas comparables sur le plan quantitatif.

De tout ceci, il résulte que les échelles d'évaluation ne permettent aucune opération arithmétique. Si on note la composition française et l'anglais deuxième langue de 0 à 10, rien ne permet d'additionner les deux résultats pour évaluer un savoir total. C'est pourtant ce que l'on fait, depuis des décennies, dans nos écoles. Il est tout aussi suspect, sinon plus, d'utiliser des échelles numériques différentes pour les différentes branches, en fonction d'une pondération empirique : de 0 à 120 pour le français, de 0 à 80 pour les sciences, etc.

Si le système a pu fonctionner, c'est qu'il a, de toute façon, permis d'identifier les meilleurs (ceux qui se classent *en tout* dans les premiers) et les plus faibles (derniers *en tout*).

C. Utilité.

Malgré ses limitations, l'échelle d'évaluation est, dans bien des cas, le seul instrument dont nous disposons pour concrétiser notre jugement sur des comportements humains complexes.

En fait, elle permet d'apporter, dans un dossier scolaire, des éléments *qualitatifs*, aussi objectifs que possible, *indispensables compléments* des résultats *quantitatifs* apportés par les examens objectifs ou les tests.

D. Construction.

Beaucoup d'éducateurs se laissent abuser par la facilité apparente avec laquelle on construit une échelle.

En réalité, il n'est pas rare que la mise au point de pareil instrument exige de longs mois de travail, nécessaires à la clarification théorique, et appelle l'utilisation de techniques raffinées comme l'analyse factorielle. Celle-ci aide notamment à déterminer dans quelle mesure plusieurs échelles évaluent les mêmes choses sous des appellations différentes.

Dans ce domaine aussi, il importe donc que, le plus rapidement possible, les enseignants puissent s'appuyer sur des services de recherche bien équipés en spécialistes et en ordinateurs.

Cela ne signifie toutefois pas que les maîtres ne peuvent rien faire en attendant.

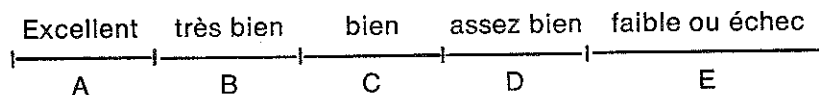
1. Combien de degrés ?

Théoriquement, une échelle d'évaluation peut compter une quantité infinie d'échelons. En pratique, les juges ne classent avec quelque sûreté que si l'on n'utilise qu'un petit nombre de degrés : 3, 5, 7 ou 9.

En nous référant à la littérature expérimentale et à de multiples observations que nous avons pu faire directement, nous conseillons que l'on s'en tienne *le plus souvent* à 5 échelons. Pareilles échelles permettent des classements relativement sûrs et fidèles, à condition que chaque échelon soit très clairement défini et les notateurs bien formés.

Pratiquement,

- 1° On classe d'abord en trois catégories : les meilleurs, les plus faibles, les autres ;
- 2° On répartit ensuite les « autres » en bons, moyens et faibles.



Remarquons que l'échelon C est à cheval sur le centre de l'échelle, c'est-à-dire sur la moyenne arithmétique. Comme nous l'avons vu, à propos de la courbe de Gauss, une forte concentration se produit souvent autour de ce point. Choisir un nombre pair d'échelons créerait un centre dans l'échelle, devant ou derrière lequel plusieurs élèves basculeraient au hasard.

Dans le cas d'une distribution normale, il n'est pas exclu que ce verdict du hasard affecte 20, 30, voire 40 % des élèves.

2. Définir l'objet de l'évaluation.

Demander, par exemple, d'apprécier le « courage » des membres d'un groupe donné — sans autre précision — conduirait à des réponses presque entièrement dépourvues d'intérêt. Quel ou quels sens les observateurs auront-ils donnés au mot « courage » : ardeur, volonté, zèle, persévérance, bravoure, fermeté, stoïcisme ?

Et même si nous précisons dans la fermeté devant le danger la distinction entre l'intrépidité et la

Pour pallier les imprécisions *bien, moyen, etc.*, il est nécessaire une description aussi précise que possible, d'apprécier, et d'illustrer la définition. Définir non par des formules mais par des exemples précis est une condition

Voici comment Schonell a défini la confiance en soi :

Confiance en soi	
Extrêmement confiant en soi. Presque trop sûr de lui.	Très confiant en ses propres forces.

Description du trait.

« Sous sa forme positive, la confiance se manifeste par les manifestations suivantes : l'enfant est capable de faire face aux difficultés, pendant et prêt à assumer des responsabilités »

» L'enfant qui a confiance en lui-même avec le minimum d'assistance doit être aidé constamment. Il est capable de construire et de suivre des instructions claires ; le second degré de confiance est celui qu'on l'aide durant toute la tâche.

Situations types.

- 1° A-t-il peur de l'obscurité ?
- 2° Peut-il prendre soin de lui-même ? faut-il que quelqu'un soit avec lui ?
- 3° Voyage-t-il seul en tramway (à partir de 9 ans).

1 F.J. SCHONELL, *Backwardness in Education: Measurement of Personality* (« Education »)

te donc que, le plus rapide-
sent s'appuyer sur des ser-
n spécialistes et en ordina-

que les maîtres ne peuvent

d'évaluation peut compter
pratique, les juges ne clas-
n n'utilise qu'un petit nom-

ure expérimentale et à de
vons pu faire directement,
e le plus souvent à 5 éche-
des classements relative-
e chaque échelon soit très
n formés.

égories : les meilleurs, les

en bons, moyens et faibles.

assez bien faible ou échec

D | E
est à cheval sur le centre de
enne arithmétique. Comme
ourbe de Gauss, une forte
tour de ce point. Choisir un
centre dans l'échelle, devant
basculeraient au hasard.

normale, il n'est pas exclu
, 30, voire 40% des élèves.

précier le « courage » des
ns autre précision — con-
èremment dépourvues d'inté-
ateurs auront-ils donnés au
ble, persévérance, bravoure,

Et même si nous précisons que par « courage » nous enten-
dons la fermeté devant le danger, ferons-nous aisément la dis-
tinction entre l'intrépidité et la témérité ?

Pour pallier les imprécisions de notations telles que *très bien, moyen, etc.*, il est nécessaire d'ajouter à l'échelle une description aussi précise que possible du trait ou de l'objet à apprécier, et d'illustrer la définition par des situations types. Définir non par des formules abstraites, mais par des comportements précis est une condition *sine qua non* de validité.

Voici comment Schonell¹ présente le trait « Confiance en soi » :

Confiance en soi

Extrêmement confiant en soi. Presque trop sûr de lui.	Très confiant en ses propres forces.	Confiant.	Manque de confiance. Timide.	Manque extrême de confiance. Dépend des autres. Décline les responsabilités.
---	--------------------------------------	-----------	------------------------------	--

Description du trait.

« Sous sa forme positive, cette qualité est marquée par les manifestations suivantes : l'individu compte sur lui-même, est capable de faire face aux difficultés, a de l'assurance, est indépendant et prêt à assumer des responsabilités.

» L'enfant qui a confiance en lui-même essaie d'avancer avec le minimum d'assistance ; celui qui manque de confiance doit être aidé constamment. Le premier aime de voir ce qu'il est capable de construire et de produire quand il a reçu des instructions claires ; le second veut qu'on lui mâche la besogne, qu'on l'aide durant toute la phase de la réalisation. »

Situations types.

- 1° A-t-il peur de l'obscurité ?
- 2° Peut-il prendre soin de lui-même et de ce qu'il possède ou faut-il que quelqu'un soit tout le temps à ses côtés ?
- 3° Voyage-t-il seul en tram ou en bus ? (pour enfants de plus de 9 ans).

¹ F.J. SCHONELL, *Backwardness in the Backward Subjects*, cité par F. WARBURNE, *Measurement of Personality* (« Educational Research », novembre 1961, p. 9).

- 4° Parle-t-il librement à des visiteurs inconnus ?
- 5° Est-il bon dans les jeux ? Sait-il nager ?
- 6° Est-il à l'aise et répond-il avec assurance aux examens oraux ?
- 7° Lit-il bien, dramatise-t-il bien un texte devant la classe ?
- 8° S'attaque-t-il bien à des tâches nouvelles ou pose-t-il constamment des questions à ses compagnons et à ses maîtres ?

Définir les degrés de l'échelle.

Exemple : Organisation de l'enseignement de la lecture.

0	1	2	3	4	5
Médiocre En lecture, tous les élèves sui- vent la même progression. Pas de travail par groupes.	Assez bien Cf. 1. Mais parfois un élève très lent reçoit un peu moins de travail que les autres.	Moyen Constitution de 2 ou 3 groupes, selon les aptitudes en lecture. Peu de flexibilité dans le groupement.	Très bien Groupement selon les aptitudes. Flexibilité.	Excellent Groupement après étude approfondie des aptitudes et des difficultés rencontrées. Grande flexibilité.	

On aura remarqué que ce dernier exemple combine les échelles graphique, numérique et descriptive.

Le danger d'une définition insuffisante des cinq degrés a été montré expérimentalement par R. Weiss¹. Adoptant les échelles descriptives imposées par le Ministère autrichien de l'Education nationale, Weiss a invité 92 instituteurs à noter deux compositions d'élèves du niveau où ils enseignaient (4^e primaire) et 272 instituteurs à noter deux travaux d'arithmétique (problèmes ; 4^e et 5^e années).

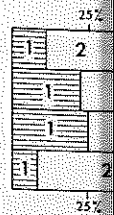
Les diagrammes ci-dessous montrent combien les notations sont divergentes, pour un même travail (dispersion des notes sur toute la longueur de l'échelle).

¹ R. WEISS, o.c.

Notation de

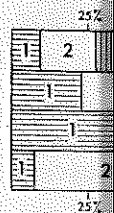
I

Orthographe
Style
Fond
Note globale



II

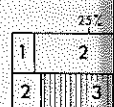
Orthographe
Style
Fond
Note globale



Notation de deux

I

II



On le voit, il ne suffit pas en remplacement du système tous les problèmes soient rés

E. Utilisation.

1) Combien d'élèves par éche

Premier cas : Elève comp

Les échelons marquent le sion vers un pouvoir ou un sa à atteindre. Dans un enseigne tif est commun à la majorité, la classe. Il doit évidemment les élèves puissent se dépa large, ce qui risque de nuire à inopérant.

inconnus ?
 ger ?
 assurance aux examens

texte devant la classe ?
 nouvelles ou pose-t-il con-
 ompagnons et à ses maî-

enseignement de la lecture.

3 4 5

Très bien	Excellent	
Groupement	Groupement	
selon les	après étude	
aptitudes.	approfondie	
Flexibilité.	des aptitudes	
	et des	
	difficultés	
	rencontrées.	
	Grande	
	flexibilité.	

rnier exemple combine les
 escriptive.

uffisante des cinq degrés a
 R. Weiss¹. Adoptant les
 le Ministère autrichien de
 92 instituteurs à noter deux
 ils enseignaient (4^e primai-
 travaux d'arithmétique (pro-

ntrent combien les notations
 travail (dispersion des notes

NOTATION DE DEUX COMPOSITIONS

Notation de deux rédactions

					Note moyenne	Sigma		
I	Orthographe	1	2	3	4	5	2.89	1.04
	Style	1	2	3	4	5	2.29	0.53
	Fond	1	2	3	4		2.08	0.79
	Note globale	1	2	3	4		2.45	0.80

					Note moyenne	Sigma		
II	Orthographe	1	2	3	4	5	2.99	0.96
	Style	1	2	3	4		2.12	1.05
	Fond	1	2	3	4		1.78	0.85
	Note globale	1	2	3	4		2.54	0.80

Notation de deux travaux d'arithmétique

					Note moyenne	Sigma
I	1	2	3	4	2.56	0.77
	2	3	4	5	3.55	0.80

On le voit, il ne suffit pas d'adopter une nouvelle échelle en remplacement du système de notes traditionnel pour que tous les problèmes soient résolus. Loin s'en faut !

E. Utilisation.

1) Combien d'élèves par échelon ?

Premier cas : Elève comparé à lui-même.

Les échelons marquent les degrés à franchir dans l'ascension vers un pouvoir ou un savoir. Le maître a défini un objectif à atteindre. Dans un enseignement non individualisé, cet objectif est commun à la majorité, sinon à la totalité des élèves de la classe. Il doit évidemment être choisi de telle sorte que tous les élèves puissent se dépasser. Ce sera donc un objectif large, ce qui risque de nuire à sa précision au point de le rendre inopérant.

Il est clair que nous agissons actuellement comme si, au début de l'année scolaire, tous les élèves se trouvaient approximativement à un même endroit de l'échelle du savoir. On suppose donc que, l'année précédente, tous ont atteint le degré supérieur de l'échelle précédente.

Or, nous le sentons bien, la réalité est différente et la recherche confirme notre sentiment :

a) *Marges de variation de l'âge mental.*

Lors de la révision du test d'intelligence de Binet-Simon par Terman et Merrill (1937), on a constaté, aux Etats-Unis¹ :

- qu'en première année primaire, l'âge mental variait de 4 à 8 ans ;
- qu'en sixième année primaire, l'âge mental allait de 8 à 16 ans ;
- que, dans l'enseignement secondaire, une marge de variation de 8 à 10 ans n'était pas exceptionnelle.

b) *Marges de variation du rendement scolaire.*

Il suffit d'observer combien les normes des tests de connaissances, étalonnés par année scolaire, se chevauchent, pour prendre conscience de l'ampleur des marges de variation.

S'appuyant sur ses propres recherches et sur celles de Lindquist, Cornell, Learned et Wood, W.-W. Cook² fournit les indications suivantes :

- Pour la compréhension de la lecture, pour le vocabulaire, les sciences, la géographie et l'histoire, la marge de variation du rendement est :
 - En 1^{re} primaire : de 3 à 4 ans.
 - En 4^e primaire : de 5 à 6 ans.
 - En 6^e primaire : de 7 à 8 ans.
- Pour le raisonnement arithmétique et le calcul :
 - En 6^e primaire : de 6 à 7 ans.
- En culture générale (sciences, littérature, arts, histoire, géographie et langue maternelle), au niveau universitaire :

¹ Cf. Q. McNEMAR, *The Revision of the Stanford-Binet Scale*, Boston, Houghton-M., 1942.

² W.W. COOK, *The Functions of Measurement in the Facilitation of Learning*, in E.F. LINDQUIST, Ed., *Educational Measurement*, Washington, A.C.E., 1961. 4^e éd. pp. 3-47.

- 28 % des étudiants de moyenne des étudiants
- 10 % des étudiants de 4^e ne des élèves de fin d'

Semblables observations pied de la lettre, mais dont il raient être multipliées à volonté

Elles prouvent qu'aussi lon neront selon le système de cla allant de pair avec le souci progresser à son allure propre aussi loin qu'il le peut, sera u

Toutefois, comme pour économiques, l'enseignement pas avant de nombreuses an individualisé offre une solutio cable. Dans une école semi-ir gènes, par branches principa Ainsi, un élève doué en langu niveau A, pour cette branche, les problèmes d'arithmétique

Afin de ne pas alourdir annexe une description détail puis cinquante ans selon ce pr

Deuxième cas : Elèves co

Même si l'on essaie d'an des différences suffisantes sub de nouveau apparaître une moins, son ébauche) où la mo cée vers le haut.

Il semble donc justifié d'e selon des pourcentages qui s' male.

¹ Ce n'est pas par hasard que nous d'arithmétique, en général. Dans une bran bien doué en calcul, mais faible en probl si l'on ne groupe pas assez finement, considérables.

actuellement comme si, au lieu de se trouver à l'échelle du savoir. On voit, tous ont atteint le degré

réalité est différente et la situation :

de mental.

intelligence de Binet-Simon constaté, aux Etats-Unis¹ :

l'âge mental variait de 4 à

l'âge mental allait de 8 à

primaire, une marge de variation exceptionnelle.

de ment scolaire.

les normes des tests de connaissance, se chevauchent, pour de larges marges de variation.

recherches et sur celles de Wood, W.-W. Cook² fournit les

lecture, pour le vocabulaire, l'histoire, la marge de varia-

que et le calcul :

mathématiques, littérature, arts, histoire, musique, (le), au niveau universitaire :

Stanford-Binet Scale, Boston, Houghton-M.,

Journal of the Facilitation of Learning, in E.F. Wood, Washington, A.C.E., 1961. 4^e éd. pp. 3-47.

- 28 % des étudiants de 4^e année sont inférieurs à la moyenne des étudiants de 2^e année ;
- 10 % des étudiants de 4^e année sont inférieurs à la moyenne des élèves de fin d'enseignement secondaire.

Semblables observations — qu'il ne faut pas prendre au pied de la lettre, mais dont il faut retenir la tendance — pourraient être multipliées à volonté.

Elles prouvent qu'aussi longtemps que nos écoles fonctionneront selon le système de classes rigides, l'évaluation continue allant de pair avec le souci de permettre à chaque élève de progresser à son allure propre, de se dépasser toujours, d'aller aussi loin qu'il le peut, sera un leurre.

Toutefois, comme pour des raisons technologiques et économiques, l'enseignement individualisé ne se généralisera pas avant de nombreuses années, seul l'enseignement semi-individualisé offre une solution réaliste, immédiatement applicable. Dans une école semi-individualisée, des groupes homogènes, par branches principales, se substituent à la classe. Ainsi, un élève doué en langue étrangère peut travailler à un niveau A, pour cette branche, mais se joindre au niveau D pour les problèmes d'arithmétique¹.

Afin de ne pas alourdir ce texte, nous avons reporté en annexe une description détaillée d'une école fonctionnant depuis cinquante ans selon ce principe.

Deuxième cas : Elèves comparés entre eux.

Même si l'on essaie d'amener chacun à un niveau élevé, des différences suffisantes subsisteront probablement pour faire de nouveau apparaître une distribution gaussienne (ou, du moins, son ébauche) où la moyenne se sera simplement déplacée vers le haut.

Il semble donc justifié d'essayer de peupler chaque échelon selon des pourcentages qui s'approchent de la distribution normale.

¹ Ce n'est pas par hasard que nous parlons de *problèmes d'arithmétique* et non d'arithmétique, en général. Dans une branche complexe comme celle-ci, un élève peut être bien doué en calcul, mais faible en problèmes. La recherche contemporaine montre que si l'on ne groupe pas assez finement, les marges de variation entre élèves restent considérables.

Dans une classe habituelle, il n'est évidemment pas possible de respecter strictement cette proportion, en raison du petit nombre d'élèves. On aura, par exemple :

La distribution normale serait :			
Excellent :	1	5 %	2,5 %
Très bons :	4	20 %	13,5 %
Bons (moyens) :	10	50 %	68 %
Faibles :	4	20 %	13,5 %
Très faible :	1	5 %	2,5 %
	<hr style="width: 100%;"/>	<hr style="width: 100%;"/>	<hr style="width: 100%;"/>
	20	100 %	100 %

Il faut y insister, se forcer à respecter pareille répartition (*distribution forcée*) n'implique pas nécessairement un échec pour un certain nombre d'élèves. Toutes les notes sont *relatives* les unes par rapport aux autres : « très faible » peut, dans certains groupes forts, se situer au-dessus de la note d'échec.

C'est d'ailleurs pourquoi *il vaut mieux d'adopter des lettres de classement* plutôt que les formules traditionnelles.

2) Lutter contre la contamination et la tendance centrale.

Outre les dangers de stéréotypie et d'effets de halo que nous décrivons, page 33, un autre phénomène de contamination, de nature plus mécanique, guette aussi les juges. S'ils utilisent consécutivement, pour un même élève, une série d'échelles orientées dans le même sens (par exemple du positif au négatif ou du meilleur au plus faible) et peut-être même présentées sur une même page, les notations tendent à se grouper dans un même côté.

On conseille de faire figurer chaque échelle sur une page différente, de tirer au sort le sens de présentation de chacune.

Enfin, il est bon de ménager un intervalle assez long après l'évaluation de chaque qualité d'un même sujet et de faire évaluer une même qualité par autant de juges que possible.

Beaucoup de juges ont aussi tendance à grouper leurs notes vers le centre de l'échelle : crainte de surévaluer ou de sous-évaluer un élève, peur de supporter la responsabilité d'un échec. La règle de la distribution forcée rend ici d'utiles services. Mais, pour être réellement efficace, elle doit pouvoir s'appuyer sur une description précise et nuancée des différents échelons.

Quand les évaluations combinées, les professeurs qui tiennent de la facilité qu'est la note comptent une large partie de leurs collègues qui osent utiliser l'ensei-

F. Comment synthétiser les

Au terme d'une période les évaluations (que l'on soustrait), faute de quoi maîtres et élèves ne peuvent pas profiter.

En revenir au simple calcul, ruiner les efforts de rigueur ordinaires, théoriquement du moins, ni soustraction.

Voici une méthode simple dans chaque branche, puis p

Au départ, on dispose de

Note	
A	
B	
C	
D	
E	
	<div style="display: flex; justify-content: space-around; width: 100%;"> + + + </div>
	Dates

Un professeur utilise au souhaite distinguer d'aspects maire, orthographe, rédaction,

Au fur et à mesure des dans le tableau, en indiquant

n'est évidemment pas possible proportion, en raison du r exemple :

La distribution normale serait :

5 %	...	2,5 %
20 %	...	13,5 %
50 %	...	68 %
20 %	...	13,5 %
5 %	...	2,5 %
<hr/>		
100 %		100 %

respecter pareille répartition pas nécessairement un échec. Toutes les notes sont relatives : « très faible » peut, dans le-dessus de la note d'échec.

ieux d'adopter des lettres de s traditionnelles.

et la tendance centrale.

typie et d'effets de halo que tre phénomène de contamination, guette aussi les juges. S'ils un même élève, une série e sens (par exemple du posi- lus faible) et peut-être même s notations tendent à se grou-

chaque échelle sur une page s de présentation de chacune.

un intervalle assez long après d'un même sujet et de faire utant de juges que possible.

ssi tendance à grouper leurs : crainte de surévaluer ou de apporter la responsabilité d'un n forcée rend ici d'utiles sert efficace, elle doit pouvoir cise et nuancée des différents

Quand les évaluations par plusieurs juges doivent être combinées, les professeurs qui se laissent séduire par la solution de facilité qu'est la note centrale doivent savoir qu'ils abdiquent une large partie de leur influence en faveur de leurs collègues qui osent utiliser l'ensemble de l'échelle.

F. Comment synthétiser les évaluations.

Au terme d'une période scolaire, il importe de synthétiser les évaluations (que l'on souhaite aussi nombreuses que possible), faute de quoi maîtres et élèves ne sauront pas les interpréter.

En revenir au simple calcul de la moyenne risquerait de ruiner les efforts de rigueur précédents. Car des évaluations ordinales, théoriquement du moins, ne permettent ni addition, ni soustraction.

Voici une méthode simple permettant de situer un élève dans chaque branche, puis pour l'ensemble des branches.

Au départ, on dispose du tableau suivant :

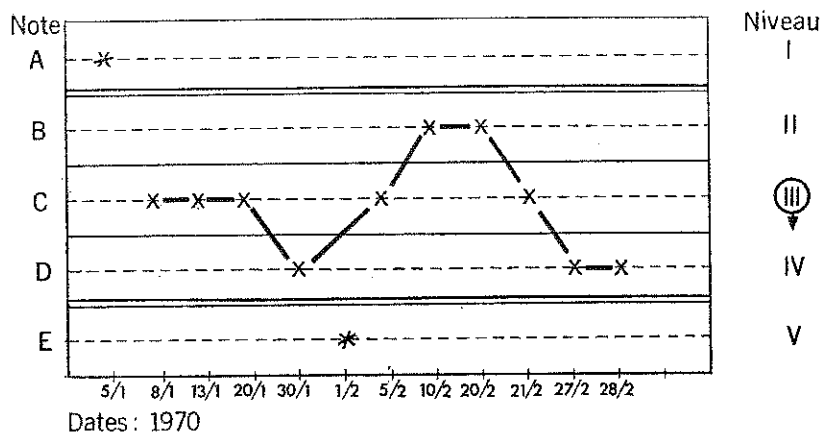
Note		Niveau
A		I
B		II
C		III
D		IV
E		V
	Dates	

Un professeur utilise autant de tableaux semblables qu'il souhaite distinguer d'aspects dans son cours : dictée, grammaire, orthographe, rédaction,...

Au fur et à mesure des évaluations, on porte une croix dans le tableau, en indiquant la date au bas.

Au moment de faire la synthèse, on commence par barrer (sans l'effacer !) la croix correspondant à la note la plus haute et celle qui représente la note la plus basse, afin d'éliminer des notes qui pourraient être accidentelles.

Supposons que l'on obtienne le tableau ci-dessous.



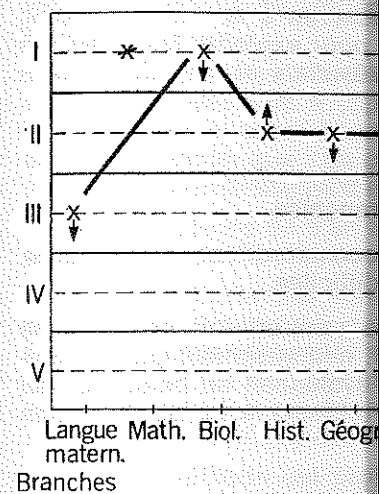
Interprétation.

- 1) Les résultats se concentrent sur trois bandes : II, III, IV. Ils ne sont donc ni excellents, ni faibles ou insuffisants. La zone de concentration est marquée par des traits doubles.
- 2) Le plus grand nombre de notes (le mode) se situent en bande III. C'est la tendance dominante.
- 3) Comme la bande la plus peuplée, après la III, est la IV, une flèche sera placée, dans cette direction, sous la III.
- 4) En réunissant les croix, on obtient un profil qui fait apparaître :
 - a) Une assez bonne homogénéité des résultats. Si l'élève se promenait anarchiquement sur les 5 bandes, il faudrait voir s'il est seul dans ce cas (cause à déterminer au niveau de l'élève), ou si la majorité de la classe se trouve dans la même situation (voir si les travaux du professeur ne sont pas mal ajustés).
 - b) Une descente en fin de période. Elle n'est pas très accusée : elle peut s'expliquer par la fatigue, par une maladie bénigne,...

On le voit, il ne suffit pas par le profil, il faut tâcher d'commentaires, oraux ou écrits et guider l'élève et ses parents.

Remarquons, par ailleurs, que si le professeur se borne à noter, sans commentaires, cela ne core moins avec une seule de certains maîtres obnubilés par l'enseignement... ou allergiques à l'usage de la note, que l'introduction d'un profil devrait heureusement améliorer.

TABLEAU POUR L'ÉVALUATION DE LA PERIODE DU



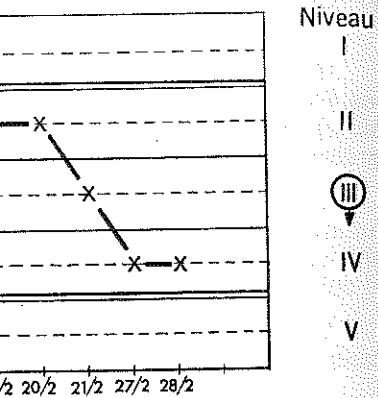
Interprétation.

Quatre notes d'ensemble sont en bande III. C'est la tendance dominante. Elle décidera de la dominante : I, II, III, IV, V.

En cas d'égalité pratique (deux notes dans deux niveaux contigus), soit deux cercles par un trait.

...se, on commence par barrer
...ndant à la note la plus haute
... plus basse, afin d'éliminer des
...elles.

le tableau ci-dessous.



...sur trois bandes : II, III, IV. Ils
...faibles ou insuffisants. La zone
...e par des traits doubles.

...notes (le mode) se situent en
...ominante.

...uplée, après la III, est la IV,
...s cette direction, sous la III.

...tient un profil qui fait apparai-

...énéité des résultats. Si l'élève
...ment sur les 5 bandes, il faudrait
...e cas (cause à déterminer au
...majorité de la classe se trouve
...oir si les travaux du professeur

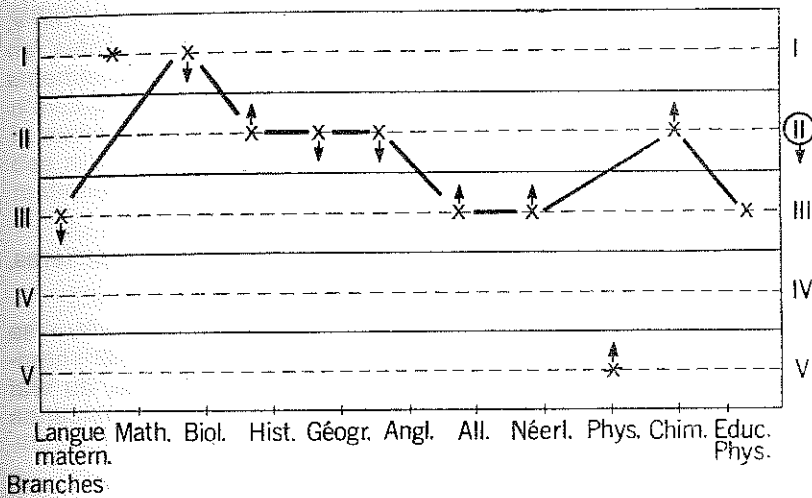
...ériode. Elle n'est pas très accu-
...r par la fatigue, par une mala-

On le voit, il ne suffit pas de constater une situation traduite par le profil, il faut tâcher d'en découvrir le pourquoi. Ainsi, les commentaires, oraux ou écrits, des maîtres viennent-ils éclairer et guider l'élève et ses parents.

Remarquons, par ailleurs, que ce système ne peut fonctionner si le professeur se borne à deux ou trois évaluations ; encore moins avec une seule ! Cette fois, c'est le comportement de certains maîtres obnubilés par la quantité de matière à enseigner... ou allergiques au long travail de notation des travaux, que l'introduction d'une nouvelle forme d'évaluation viendra heureusement améliorer.

TABEAU POUR L'ENSEMBLE DES BRANCHES

PERIODE DU AU 19 ..



Interprétation.

Quatre notes d'ensemble se trouvent en bande II et quatre en bande III. C'est la tendance du profil (ici vers le haut) qui décidera de la dominante : ici II, avec flèche vers le bas.

En cas d'égalité pratique entre deux niveaux, on peut soit décider que le niveau supérieur l'emporte (surtout s'il s'agit de deux niveaux contigus), soit entourer deux chiffres et réunir les deux cercles par un trait.

Remarques.

1° Ce système de synthèse des évaluations ne détient aucun pouvoir magique ! Si l'on commet des erreurs signalées dans le présent chapitre, la note de synthèse sera probablement dépourvue de sens.

Un phénomène de concentration dans les bandes médianes étonne souvent les utilisateurs. Ils oublient que, dans le système traditionnel, aboutissant à un pourcentage, les notes se concentraient aussi dans la même zone : entre 65 et 75 ou 80 %, tout le reste de l'échelle (de 100 échelons !) étant bien souvent presque vide.

En réalité, le phénomène de concentration est plus frappant parce que l'échelle à cinq degrés est beaucoup mieux visualisée que l'ancienne.

2° Rien n'empêche de pondérer les notes, c'est-à-dire d'accorder, par exemple, deux fois plus d'importance au travail en classe qu'aux travaux à domicile, à tel aspect d'une branche plutôt qu'à tel autre, etc.

3° On trouvera, dans la partie consacrée à la « modération », des indications supplémentaires pour exploiter les notes, en particulier pour déterminer les résultats et les classements en fin d'année (voir page 171).

G. Un cas particulier :

la notation de la composition française.

Ce problème mérite quelques considérations particulières, parce qu'il compte parmi les plus difficiles.

S'il était résolu, la docimologie remporterait peut-être sa plus belle victoire. Hélas ! ou, plutôt, heureusement pour l'homme, la plus noble de ses activités, l'évaluation du beau, du vrai et du bien, échappera sans doute toujours à la quantification objective et donc automatisable.

1) Quatre méthodes d'évaluation.

Les principales méthodes se ramènent à quatre : la méthode de l'impression générale, l'échelle de spécimens, la méthode analytique et les comptages de fréquences.

a) La méthode de l'impression générale.

C'est la plus subjective. On lui reproche le manque de fidélité d'un même correcteur, les désaccords entre plusieurs notateurs, spécialistes ou non, et son inefficacité fréquente : maint professeur s'est usé à noter avec soin des centaines de travaux par an sans obtenir des résultats convaincants.

Il y aurait beaucoup à dire sur la position française, et surtout sur la profondeur des habitudes scolaires. Comme un écrivain en puissance, on ne peut pas d'en traiter. Retenons simplement que deux faiblesses souvent accusées, la méthodologie, grève l'exercice que tellement lourde que ce n'est pas de la notation. Nous ne sommes pas de ceux qui...

Malgré ses faiblesses, la notation est la mieux en harmonie avec la composition. En utilisant un maximum, à cinq degrés, on obtient un résultat obtenu entre des notateurs et appartenant à un même niveau. Il faut donc prendre les précautions à prendre.

Une composition corrigée, croyons-nous, ambitionner d'être plus qu'une simple correction du langage à cet effet.

La composition traditionnelle d'exception, destinée à repérer la maturation, et aussi à identifier les maladresses que l'on corrige par des exercices méthodiques d'enrichissement et de perfectionnement.

Si l'on décide de s'en tenir à la technique de l'évaluation traditionnelle...

Elle comporte des variantes. Celle proposée par S. Wiseman¹ est la plus intéressante. Elle évalue indépendamment des compositions une allure d'environ cinquante pour les obliger à se décider sur une moyenne arithmétique des notes.

Cette méthode est particulièrement adaptée aux travaux narratifs ou descriptifs recouvrant toutes les compositions. Elle atteste la supériorité de la méthode traditionnelle, pourtant valide aussi en ce qui concerne le haut niveau intellectuel.

La validité générale de la « dissertation » est confirmée par les statistiques de J. Britton.

¹ S. WISEMAN, The Marking of English Compositions, *British Journal of Educational Psychology*, 1954, 24, 1-10.

valuations ne détient aucun
nmet des erreurs signalées
note de synthèse sera pro-

n dans les bandes médianes
ls oublient que, dans le sys-
a un pourcentage, les notes
même zone : entre 65 et 75
ille (de 100 échelons !) étant

ncentration est plus frappant
s est beaucoup mieux visua-

s notes, c'est-à-dire d'accor-
s d'importance au travail en
e, à tel aspect d'une branche

nsacrée à la « modération »,
pour exploiter les notes, en
résultats et les classements

française.

considérations particulières,
difficiles.

ie remporterait peut-être sa
ot, heureusement pour l'hom-
l'évaluation du beau, du vrai
toujours à la quantification

amèment à quatre : la métho-
le de spécimens, la méthode
uences.

n générale.

reproche le manque de fidé-
désaccords entre plusieurs
son inefficacité fréquente :
avec soin des centaines de
résultats convaincants.

Il y aurait beaucoup à dire sur la méthodologie de la composition française, et surtout sur la nécessité d'une réforme profonde des habitudes scolaires qui traitent chaque élève comme un écrivain en puissance, mais il ne nous appartient pas d'en traiter. Retenons simplement que la conjugaison de deux faiblesses souvent accusées, l'une docimologique, l'autre méthodologique, grève l'exercice de rédaction d'une hypothèse tellement lourde que certains déclarent cet exercice inutile. Nous ne sommes pas de cet avis.

Malgré ses faiblesses, la méthode de l'impression générale est la mieux en harmonie avec la complexité essentielle de la composition. En utilisant une échelle courte, à trois ou, au maximum, à cinq degrés, un accord assez élevé peut être obtenu entre des notateurs expérimentés, de même formation et appartenant à un même milieu scolaire. Nous verrons plus loin les précautions à prendre avant de décider d'un échec.

Une composition corrigée par cette méthode ne doit pas, croyons-nous, ambitionner d'améliorer beaucoup le style ou la simple correction du langage écrit : d'autres exercices existent à cet effet.

La composition traditionnelle doit être une prestation d'exception, destinée à repérer des talents ou à en contrôler la maturation, et aussi à identifier des faiblesses, des carences, des maladroites que l'on combattra dans des exercices systématiques d'enrichissement et de remédiation.

Si l'on décide de s'en tenir à la seule appréciation globale, la technique de l'évaluation en temps imposé mérite l'attention.

Elle comporte des variantes mineures. La méthode proposée par S. Wiseman¹ est caractéristique. Quatre notateurs évaluent indépendamment des travaux faits en trente minutes, à une allure d'environ cinquante copies à l'heure (taux imposé pour les obliger à se décider très rapidement). On calcule la moyenne arithmétique des notes.

Cette méthode est probablement la plus efficace pour les travaux narratifs ou descriptifs. Pour les « dissertations », mot recouvrant toutes les compositions où le sujet doit argumenter, la supériorité de la méthode est plus discutée. Elle semble pourtant valide aussi en ce domaine si les notateurs sont de haut niveau intellectuel.

La validité générale de la méthode et son applicabilité à la « dissertation » est confirmée par deux recherches expérimentales de J. Britton.

1 S. WISEMAN, The Marking of English Composition in Grammar School Selection, in *British Journal of Educational Psychology*, XIX, nov. 1949, 208.

1^{re} expérience ¹.

Un ensemble de compositions écrites par 168 élèves de 15 ans a d'abord été noté, selon la méthode analytique, par un examinateur expérimenté. Les mêmes travaux ont été ensuite évalués indépendamment, selon la méthode de l'impression générale en temps imposé, par huit notateurs.

Pour disposer d'un critère, cinq autres travaux de chaque élève ont été notés par deux examinateurs.

Les notes analytiques et la note moyenne de trois impressions rapides tirées au hasard parmi les huit dont on disposait pour chaque étudiant, ont alors été comparées aux notes de critère :

corrélation entre notes analytiques et critère : . 71

corrélation entre notes rapides et critère : . 76

En outre, la fidélité de la notation rapide a été calculée en comparant les notes de deux groupes de trois examinateurs, choisis au hasard parmi les huit. On obtient un coefficient très élevé : . 87.

2^e expérience ².

Cette seconde expérience, beaucoup plus fine que la précédente, a porté sur un échantillon de 500 élèves de 17 ans, stratifié selon le sexe, le type d'école secondaire et le lieu d'habitation (Londres, centre et faubourgs, grandes villes, petites villes). Tous les types de sujets de compositions, de la simple narration à la dissertation philosophique et à l'amplification poétique, étaient aussi représentés.

De nouveau, la notation multiple selon l'impression rapide s'est montrée plus fidèle et plus valide que la notation analytique.

b) Les échelles de spécimens.

Un petit nombre de compositions, souvent cinq, sont choisies de façon à constituer des modèles représentatifs des différents niveaux de qualité : de la médiocrité à l'excellence.

Le notateur évalue les travaux par comparaison aux cinq textes-étalons de l'échelle.

¹ J. BRITTON, *Experimental Marking of English Composition Written by Fifteen-Year-Olds*, in *Educational Review* (Birmingham), vol. 16, 1, 1963, 17-23.

² J. BRITTON, N. MARTIN et H. ROSEN, *Multiple Marking of English Composition*, Londres, H.M.S.O., 1966.

Après avoir connu une v
pays ¹, les échelles de spéci
plusieurs raisons ² :

1. Il est rare que la composition aux textes-étalons ;
2. Les notateurs ont tendances communes qui les influencent les autres ;
3. Des échelles différentes s
sujet et selon le niveau des

D. Pidgeon nous a signalé récemment en Grande-Bretagne à 5 degrés, s'est soldée par un correcteurs. Le résultat n'est que des impressions purement subjectives.

Notons cependant que, dans des échelles de spécimens réalisés par des chercheurs.

c) La méthode analytique

Deux ou trois notateurs sont choisis pour chacune des différentes échelles d'observer. On totalise et on compare.

Cette méthode est la plus souvent critiquée parce qu'elle ne permet pas d'échapper à toute discussion. On remarque d'ailleurs que, dans le travail estimé unanimement par les experts, se classe rarement en

Pourtant, la recherche montre que l'importance à leur accord. La méthode analytique est la plus précise. Le seul correcteur. La méthode multiple nous l'avons déjà vu, d'ailleurs, dans la participation de plusieurs notateurs.

¹ Pour une étude d'ensemble, voir : J. BRITTON, *Rhetoric Found in English Compositions*, N

² R. BRADDOCK et al., *Research in Education*, 1963, p. 12.

³ On trouvera aussi une bonne étude sur les différentes méthodes de notation de l'anglais dans *Journal of Educational Psychology*, IX, Nov. 1939, 257-269 et X, Feb.

écrites par 168 élèves de méthode analytique, par un mes travaux ont été ensuite la méthode de l'impression notateurs.

q autres travaux de chaque ateurs.

te moyenne de trois impres- ni les huit dont on disposait é comparées aux notes de

es et critère : .71

et critère : .76

ion rapide a été calculée en upes de trois examinateurs, n obtient un coefficient très

ucoup plus fine que la pré- n de 500 élèves de 17 ans, école secondaire et le lieu bourgs, grandes villes, peti- ets de compositions, de la bphilosophique et à l'amplifi- sentés.

le selon l'impression rapide alide que la notation analy-

ns, souvent cinq, sont choi- èles représentatifs des diffé- édiocrité à l'excellence.

par comparaison aux cinq

sh Composition Written by Fifteen-Year- 16, 1, 1963, 17-23.

ultiple Marking of English Composition,

Après avoir connu une vogue considérable dans certains pays¹, les échelles de spécimens ont été abandonnées pour plusieurs raisons² :

1. Il est rare que la composition à noter ressemble étroitement aux textes-étalons ;
2. Les notateurs ont tendance à n'identifier que les caractéristiques communes qui les intéressent spécialement et négligent les autres ;
3. Des échelles différentes sont nécessaires selon le type de sujet et selon le niveau des élèves.

D. Pidgeon nous a signalé qu'une vaste expérience, menée récemment en Grande-Bretagne, avec une échelle de spécimens à 5 degrés, s'est soldée par un écart moyen de 2 degrés entre correcteurs. Le résultat n'est donc guère meilleur qu'en conditions purement subjectives.

Notons cependant que, depuis 1903 au moins, la méthode des échelles de spécimens retrouve périodiquement la faveur des chercheurs.

c) La méthode analytique.

Deux ou trois notateurs accordent un certain nombre de points pour chacune des différentes qualités qu'il est convenu d'observer. On totalise et on calcule la moyenne entre notateurs.

Cette méthode est la plus lourde des quatre ; elle est souvent critiquée parce qu'elle dissèque un tout qui semble précisément échapper à toute dissection systématique. S. Wiseman remarque d'ailleurs que, dans un ensemble noté analytiquement, le travail estimé unanimement comme le meilleur, par les experts, se classe rarement en tête.

Pourtant, la recherche montre que, si les qualités à observer et l'importance à leur accorder sont définies avec précision, la méthode analytique est la plus sûre, si l'on doit se fier à un seul correcteur. La méthode de l'impression générale donne, nous l'avons déjà vu, d'aussi bons résultats, mais elle exige la participation de plusieurs notateurs³.

¹ Pour une étude d'ensemble, voir : E. HINTON, *Study of the Qualities of Style and Rhetoric Found in English Compositions*, New York, 1940.

² R. BRADDOCK et al., *Research in Written Composition*, Champaign, Ill, N.C.T.E., 1963, p. 12.

³ On trouvera aussi une bonne étude comparative in B. CAST, *The Efficiency of Different Methods of Marking English Composition*, in *British Journal of Educational Psychology*, IX, Nov. 1939, 257-269 et X, Feb. 1940, 49-60, cité par R. BRADDOCK, o.c., p. 13.

Pour être efficace et praticable, la méthode analytique semble devoir répondre à deux exigences au moins :

- 1° Les correcteurs doivent accepter le plan d'analyse. L'idéal est qu'ils participent à son élaboration ;
- 2° Le plan ne doit pas être trop détaillé. Ainsi, la méthode très fine utilisée par E. Burton¹ n'offre d'intérêt que pour la recherche, en raison du temps très long qu'elle exige.

Quelles qualités observer ?

C. Remondino² a montré que les attitudes fondamentales des correcteurs cultivés (professeurs ou non) ne diffèrent pas essentiellement : les mêmes qualités retiennent leur attention. Mais ils leur accordent des poids très variables, dans l'appréciation d'ensemble, et là gît réellement la source des divergences.

Elaborer un tableau très détaillé d'analyse du contenu est un leurre, car, même si les nuances distinguées existent, les notateurs les refondent inconsciemment. L'analyse factorielle met ce processus en évidence. Le travail de Remondino apporte d'intéressants résultats à ce propos.

Remondino a d'abord interrogé longuement vingt professeurs, de branches littéraires (enseignement secondaire) et il a aussi dressé la liste des qualités relevées dans des compositions scolaires. Il aboutit à l'ensemble suivant :

1. Lisibilité
2. Esthétique
3. Présentation
4. Exactitude de l'orthographe
5. Exactitude morphologique
6. Exactitude syntaxique
7. Structure de l'exposé
8. Richesse d'idées
9. Pertinence des idées
10. Précision d'information
11. Exhaustivité
12. Concision
13. Propriété du langage
14. Style
15. Originalité
16. Maturité
17. Imagination

Une analyse factorielle
lités relevaient de quatre

1. Présentation graphique
2. Usage de la langue
3. Contenu et organisation
4. Aspects personnels

¹ E. BUXTON, An Experiment to Test the Effects of Writing Frequency..., in *Alberta Journal of Educational Research*, V, Juin 1959, 91-99.

² Etude factorielle sur la notation des compositions scolaires portant sur la langue maternelle, in *Le Travail Humain*, XXII, Janv.-Juin 1959, 27-40.

le, la méthode analytique
gences au moins :

le plan d'analyse. L'idéal
ation ;

taillé. Ainsi, la méthode très
offre d'intérêt que pour la
très long qu'elle exige.

es attitudes fondamentales
s ou non) ne diffèrent pas
s retiennent leur attention.
ès variables, dans l'appré-
ent la source des divergen-

é d'analyse du contenu est
s distinguées existent, les
ment. L'analyse factorielle
avail de Remondino apporte

longuement vingt profes-
gnement secondaire) et il a
élevées dans des composi-
ble suivant :

1. Lisibilité	Qualité d'une écriture qui se prête à une lecture facile, rapide, sans équivoque.
2. Esthétique	Ligne harmonieuse et agréable des lettres et bon goût dans la mise en pages.
3. Présentation	Propreté, soin, bonne présentation de la copie.
4. Exactitude de l'orthographe	Densité des erreurs d'orthographe.
5. Exactitude morphologique	Densité des erreurs morphologiques.
6. Exactitude syntaxique	Densité des erreurs de syntaxe.
7. Structure de l'exposé	Qualité d'un exposé fait avec ordre, dans les proportions voulues, et selon un plan.
8. Richesse d'idées	Quantité d'idées ; ressources utilisées.
9. Pertinence des idées	Qualité par laquelle les idées exposées sont en juste rapport avec le thème traité.
10. Précision d'information	Véracité et exactitude objective des affirmations et des faits exposés.
11. Exhaustivité	Qualité qui consiste à ne rien laisser de côté de tout ce qui devait se dire.
12. Concision	Qualité par laquelle les choses à dire le sont avec le minimum indispensable de termes sans répétitions, redondance ou tortuosité.
13. Propriété du langage	Juste emploi des termes.
14. Style	Facilité, exactitude, maîtrise de la langue au point de vue de la « construction des phrases ».
15. Originalité	Qualité par laquelle, à travers le travail, transparait et s'affirme quelque chose de la personnalité.
16. Maturité	Capacité de jugement, profondeur critique, acuité des raisonnements exposés.
17. Imagination	Capacité de création, de transfiguration, de « projection » révélée par le travail.

Une analyse factorielle a ensuite montré que ces 17 qualités relevaient de quatre groupes seulement :

1. Présentation graphique (1, 2 et 3) ;
2. Usage de la langue (4, 5, 6, 13, 14) ;
3. Contenu et organisation de l'exposé (7, 8, 9, 10, 11) ;
4. Aspects personnels du fond (15, 16 et 17).

acts of Writing Frequency..., in *Alberta*
ositions scolaires portant sur la langue
9, 27-40.

Un exemple.

Les conclusions de Remondino recourent largement celles des autres chercheurs. Les échelles suivantes, construites par l'*Educational Testing Service* (E.T.S.) de Princeton¹, et légèrement remaniées par E. Page², témoignent de la similitude des vues.

Nous traduisons le texte complet parce qu'il apporte un des meilleurs exemples que l'on puisse donner actuellement pour indiquer la voie à suivre.

CRITERES POUR NOTER LES COMPOSITIONS.

I. Définition des traits à évaluer.

- A. Idées ou contenu : la quantité et la qualité du matériel utilisé pour traiter du sujet.
- B. Organisation : la relation entre les parties de la composition et l'ensemble.
- C. Style : utilisation du langage au-delà de la simple correction grammaticale.
- D. Mécanique : orthographe, grammaire, ponctuation.
- E. Créativité.

II. Guide pour l'évaluation de ces cinq traits.

A. Idées ou contenu.

Niveau élevé.

L'étudiant traite de tous les points appelés par le sujet ou le plan de travail. Il comprend bien le sujet et utilise des définitions claires. Il sait considérer le sujet dans une perspective plus large que celle des autres élèves de la classe. Autrement dit, il témoigne d'une expérience plus riche.

Niveau moyen.

Les idées sont appropriées, mais conventionnelles et peu nombreuses. Certains aspects du sujet sont négligés. L'élève ne semble pas avoir un esprit richement meublé.

¹ E.T.S., *Definitions of Ratings on the E.T.S. Compositions Scale*, cité par E. PAGE, o.c., pp. 76-77.

² O.c., pp. 78-80.

Niveau bas.

L'étudiant omet beaucoup. Il semble ne pas disposer d'informations au sujet et, par conséquent, idées simples.

B. Organisation.

Niveau élevé.

L'étudiant suit un plan de travail. Il avance des raisons. S'il décrit quelque chose, il va du sommet à la base, par ordre de complexité, etc.). Si l'étudiant explique, il utilise un plan cohérent d'analyse. L'étudiant sent bien ce qu'il dit. Il évite les répétitions. Il témoigne de la relation entre les différentes parties de son travail.

Niveau moyen.

L'étudiant ne s'en tient pas à un plan. Ses idées sans rapport avec le sujet. Il traite des choses peu importantes ou il y a une association libre (qu'est-ce que ça a à voir avec l'Hawaï ?) plutôt qu'en

Niveau bas.

L'étudiant ne semble pas s'en tenir à un plan avant de commencer à écrire. Il prend une direction, puis en change. Il y a un core, jusqu'à ce que le lecteur s'égare. Les idées ne sont pas clairement séparées. L'ordre de présentation est laissé au hasard.

C. Style.

(Plusieurs aspects du style sont évalués : individualité, vivacité, originalité, nous intéressons ici à trois aspects : répétition, variation et éventail des ressources.)

Niveau élevé.

L'étudiant utilise un langage riche et varié. La présentation du travail. Il utilise des ressources habituelles. Les mots sont présentés de manière intéressante.

recoupent largement celles
suivantes, construites par
(S.) de Princeton¹, et légè-
rement de la similitude des

plet parce qu'il apporte un
puisse donner actuellement

ES COMPOSITIONS.

la qualité du matériel utilisé

es parties de la composition

delà de la simple correction

aire, ponctuation.

q traits.

nts appelés par le sujet ou
e sujet et utilise des défini-
ujet dans une perspective
es de la classe. Autrement
s riche.

is conventionnelles et peu
et sont négligés. L'élève ne
nt meublé.

ompositions Scale, cité par E PAGE,

Niveau bas.

L'étudiant omet beaucoup d'aspects importants du sujet. Il semble ne pas disposer d'une réserve de connaissances relatives au sujet et, par conséquent, répète sans cesse quelques idées simples.

B. Organisation.

Niveau élevé.

L'étudiant suit un plan défini. S'il présente le pour et le contre, il avance des raisons pertinentes, dans un ordre efficace. S'il décrit quelque chose, il le fait de façon ordonnée (du sommet à la base, par ordre d'importance, par ordre de complexité, etc.). Si l'étudiant explique un concept ou un processus, il utilise un plan cohérent d'analyse, de définition ou d'illustration. L'étudiant sent bien ce qui se rapporte à son plan et évite des répétitions. Il témoigne du sens de la mesure en traitant les différentes parties de son travail.

Niveau moyen.

L'étudiant ne s'en tient pas à son plan ou introduit des idées sans rapport avec le sujet. Il consacre trop de temps à des choses peu importantes ou se répète. Il traite le sujet par association libre (qu'est-ce qui me vient à l'esprit quand je pense à Hawaï ?) plutôt qu'en poursuivant un but bien défini.

Niveau bas.

L'étudiant ne semble pas s'être demandé ce qu'il allait faire avant de commencer à écrire. Il ne suit pas de plan. Le travail prend une direction, puis en change, en change encore et encore, jusqu'à ce que le lecteur soit perdu. Les points principaux ne sont pas clairement séparés les uns des autres, et leur ordre de présentation est laissé au hasard.

C. Style.

(Plusieurs aspects du style peuvent intervenir dans l'évaluation : individualité, vivacité, élégance, etc. Toutefois, nous nous intéressons ici à trois aspects stylistiques seulement : clarté, variation et éventail des ressources linguistiques.)

Niveau élevé.

L'étudiant utilise un langage qui rend aisé la compréhension du travail. Il utilise des mots adéquats, dans leur sens habituel. Les mots sont présentés dans un ordre normal. Les

transitions sont bien ménagées. L'élève évite les ambiguïtés et ne trompe pas l'attente du lecteur. En même temps, l'étudiant évite la répétition monotone de mots, de compléments ou de structures de phrases. Finalement, il témoigne de la connaissance d'un large éventail de ressources linguistiques. Son vocabulaire est bon. Il utilise des structures parallèles ou fait un usage subtil de la subordination.

Niveau moyen.

L'étudiant égare parfois le lecteur en utilisant un mot inapproprié ou une tournure bizarre ; ou bien en utilisant une métaphore peu claire, ou en déplaçant de façon inopportune un complément ou une subordonnée, ou encore en pratiquant des transitions abruptes. La répétition de mots, de tournures et de structures de phrases devient monotone. Les ressources linguistiques sont limitées. L'élève utilise volontiers des clichés et des tournures éculées.

Niveau bas.

Les mots sont utilisés de façon vague. Tournures ambiguës, constructions boiteuses, vocabulaire et structures de phrases enfantins.

D. Mécanique.

Niveau élevé.

La structure des phrases est habituellement correcte, même lorsqu'il s'agit de modèles variés et compliqués. Les règles de l'orthographe sont respectées, même les mots difficiles sont généralement écrits sans faute. Pas de violation grave des règles de ponctuation, de majuscules, d'abréviations, d'écriture des nombres.

Niveau moyen.

Défauts de syntaxe occasionnels. Les mots difficiles sont parfois mal orthographiés. Quelques violations des règles de ponctuation, etc.

Niveau bas.

Très grand nombre de fautes.

E. Créativité.

Niveau élevé.

L'étudiant surprend par d considérer le problème. Il int son traitement du sujet. Il trou santes d'utiliser le langage po

Niveau moyen.

L'étudiant pense à ce que traite les choses comme à peu expressions et des structures

Niveau bas.

L'étudiant utilise des clic traite le sujet de façon superf réellement en comprendre la

d) La méthode des comp

Nous la signalons ici, bi la catégorie des évaluations tableau des méthodes.

Elle consiste à définir (étude diagnostique), à compt à exprimer le nombre obtenu

Au lieu d'adopter un po faire l'inventaire (relativemen utilisés dans la composition, méthode est fourni par les *El lyses littéraires* de A. Purves¹

Une dernière méthode aspects qui se révèlent préd normalement attribuée par c nateur, programmé en conse la notation ; c'est ce qu'a fa

Les comptages de fréqu espoir de notation objectif toutefois devoir rester du do la pratique scolaire, à moins

¹ A. PURVES et V. RIPPERE, *Elé* (traduit de l'anglais par J. Dubois), in 229-334.

E. Créativité.

Niveau élevé.

L'étudiant surprend par des façons neuves et efficaces de considérer le problème. Il introduit des idées nouvelles dans son traitement du sujet. Il trouve des façons fraîches et intéressantes d'utiliser le langage pour faire ressortir ses idées.

Niveau moyen.

L'étudiant pense à ce que l'on s'attendait qu'il penserait. Il traite les choses comme à peu près tout le monde. Il utilise des expressions et des structures de phrases ordinaires.

Niveau bas.

L'étudiant utilise des clichés de pensée et d'expression. Il traite le sujet de façon superficielle. Il répète des formules sans réellement en comprendre la signification.

d) La méthode des comptages de fréquences.

Nous la signalons ici, bien qu'elle n'ait pas de place dans la catégorie des évaluations subjectives, afin de compléter le tableau des méthodes.

Elle consiste à définir des types de fautes (après une étude diagnostique), à compter celles-ci dans la composition et à exprimer le nombre obtenu pour 100 ou pour 1 000 mots.

Au lieu d'adopter un point de vue négatif, on peut aussi faire l'inventaire (relativement exhaustif) des divers processus utilisés dans la composition. Le plus bel exemple de pareille méthode est fourni par les *Eléments pour servir à l'étude d'analyses littéraires* de A. Purves¹.

Une dernière méthode positive consiste à isoler certains aspects qui se révèlent prédictifs, à un seuil donné, de la note normalement attribuée par des correcteurs qualifiés. Un ordinateur, programmé en conséquence, peut alors se charger de la notation ; c'est ce qu'a fait E. Page.

Les comptages de fréquences offrent probablement le seul espoir de notation objective des compositions. Ils semblent toutefois devoir rester du domaine de la recherche étrangère à la pratique scolaire, à moins que l'ordinateur ne devienne d'un

¹ A. PURVES et V. RIPPERE, *Eléments pour servir à l'étude d'analyses littéraires* (traduit de l'anglais par J. Dubois), in *Scientia Paedagogica Experimentalis*, VI, 2, 1969, 229-334.

accès tellement facile que l'étude des fréquences se fasse de façon économique et constitue un des aspects de l'évaluation. Ceci n'est nullement exclu.

2) Plusieurs sujets au choix ?

J. Britton remarque avec raison que la rédaction d'examen sur un sujet unique implique souvent, chez le professeur, l'hypothèse qu'un sujet comme « Nuages » révélera aussi bien l'aptitude d'un étudiant que « Exposez vos vues sur la peine capitale ».

On a montré expérimentalement qu'il n'en est rien, nous allons y revenir. Toutefois, il est douteux que la justice soit beaucoup mieux respectée si l'on offre trois sujets au choix quand les points d'examens sont attribués pour une rédaction unique. Il faudrait d'abord prouver qu'avec trois, voire cinq sujets, on recouvre tout le spectre des stimuli nécessaires pour donner à chacun une chance égale ; et même si cela était établi, il importerait encore que le professeur sélectionne toujours les sujets à bon escient.

L'expérience, disions-nous, nous montre que le rendement de l'élève varie selon le thème proposé. Finlayson¹ a établi, pour la fin de l'école primaire, que si, à une semaine d'intervalle, les élèves sont invités à faire une rédaction à partir d'un même choix de quatre sujets, ceux qui changent de sujets (plus de la moitié dans l'expérience) obtiennent des notes significativement différentes.

Wiseman et Wrigley² arrivent à la même conclusion (choix parmi le même ensemble de cinq sujets à quatre mois d'intervalle), tout en prouvant que la différence de note n'est pas imputable au manque de fidélité du professeur. Vernon et Millican³ confirment la conclusion au niveau universitaire.

Que faire ? Si l'on veut vérifier la capacité de s'exprimer par écrit dans une perspective étroitement définie (par exemple : capacité de faire un rapport scientifique ou d'écrire une lettre commerciale), la sensibilité particulière des candidats n'a guère à être considérée et un sujet suffit.

Par contre, s'il s'agit ou bien de déceler une aptitude à l'expression, où qu'elle soit, ou d'amener les élèves à discuter de problèmes où la tournure d'esprit et le niveau d'information

1 D.S. FINLAYSON, The Reliability of the Marking of Essays, in *British Journal of Educational Psychology*, XXI, 2, 1951, cité par Britton (1966).

2 S. WISEMAN et J. WRIGLEY, Essay-Reliability: the Effect of Choice of Essay-Title, in *Educational and Psychological Measurement*, 18, 1, 1958.

3 P. VERNON et G. MILLICAN, A Further Study of the Reliability of English Essays, in *British Journal of Statistical Psychology*, VII, 2, 1954.

peuvent jouer un grand rôle, il jugement à partir d'une seule se référer aussi au travail de pu choisir parmi plusieurs su réagi à plusieurs dizaines de mois et l'on peut supposer chance. Laisser inventer le l'on pouvait se prémunir co C'est bien difficile...

3) Conclusion.

Le sort à réserver à la co but poursuivi. Que veut-on ? Dans ce cas, il vaut mieux n travaux ou l'attaquer dans de

Connaître la capacité des la prose conforme au bon usag avec réponses à choix multipl

Faire acquérir la capacité ce cas, des thèmes très limités tion sur l'objet du travail. Pou « Décrivez un vélo de course »

Déceler des talents d'écri cours à l'expression de l'élève même son sujet.

Mais que l'on ne s'y tro capacités différentes. Dans la s décrite plus haut, l'auteur a mo

1° Que la corrélation entre les et un résumé de texte varie

2° Que la corrélation de la ré hension de texte est un peu

Les écarts sont donc consi

Nous l'avons déjà dit, nous nir la composition traditionnelle avait été supprimée à cause de reuse, on l'a d'ailleurs réintro où la Commission des Examen ment Supérieur a renoncé au « en Angleterre, aux deux épreuv le 16 + Examenation.

des fréquences se fasse de
des aspects de l'évaluation.

que la rédaction d'examen
t, chez le professeur, l'hypo-
» révélera aussi bien l'apti-
vos vues sur la peine capi-

nt qu'il n'en est rien, nous
douteux que la justice soit
offre trois sujets au choix
attribués pour une rédaction
r qu'avec trois, voire cinq
des stimuli nécessaires pour
et même si cela était établi,
leur sélectionne toujours les

as montre que le rendement
oposé. Finlayson¹ a établi,
si, à une semaine d'inter-
une rédaction à partir d'un
qui changent de sujets (plus
ennent des notes significati-

la même conclusion (choix
sujets à quatre mois d'inter-
fférence de note n'est pas
professeur. Vernon et Malli-
niveau universitaire.

r la capacité de s'exprimer
bitement définie (par exem-
scientifique ou d'écrire une
articulière des candidats n'a
uffit.

de déceler une aptitude à
mener les élèves à discuter
it et le niveau d'information

arking of Essays, in *British Journal of*
(1966).
ility: the Effect of Choice of Essay-
t, 18, 1, 1958.
dy of the Reliability of English Essays,
354.

peuvent jouer un grand rôle, il n'est guère possible de porter un jugement à partir d'une seule composition. L'idéal est alors de se référer aussi au travail de l'année ; si, chaque fois, l'élève a pu choisir parmi plusieurs sujets (cinq, par exemple), il aura réagi à plusieurs dizaines de thèmes différents en quelques mois et l'on peut supposer que chacun aura vraiment eu sa chance. Laisser inventer le sujet à traiter serait tentant, si l'on pouvait se prémunir contre la préparation frauduleuse. C'est bien difficile...

3) Conclusion.

Le sort à réserver à la composition française dépendra du but poursuivi. Que veut-on ? Corriger un défaut particulier ? Dans ce cas, il vaut mieux n'examiner que celui-là dans les travaux ou l'attaquer dans des exercices spéciaux.

Connaître la capacité des élèves à se corriger, à identifier la prose conforme au bon usage ? Dans ce cas, un test objectif, avec réponses à choix multiple, peut faire l'affaire.

Faire acquérir la capacité de s'exprimer clairement ? Dans ce cas, des thèmes très limités permettent de concentrer l'attention sur l'objet du travail. Pourquoi bannir les sujets tels que : « Décrivez un vélo de course » ?

Déceler des talents d'écrivains ? Alors, il faut laisser libre cours à l'expression de l'élève, l'invitant souvent à choisir lui-même son sujet.

Mais que l'on ne s'y trompe pas, on testera ainsi des capacités différentes. Dans la seconde expérience de J. Britton, décrite plus haut, l'auteur a montré :

- 1° Que la corrélation entre les résultats à la rédaction créative et un résumé de texte varie de .30 à .40 ;
- 2° Que la corrélation de la rédaction avec un test de compréhension de texte est un peu plus élevée : de .35 à .45.

Les écarts sont donc considérables.

Nous l'avons déjà dit, nous ne pensons pas qu'il faille bannir la composition traditionnelle. Dans certains pays où elle avait été supprimée à cause de la difficulté d'évaluation rigoureuse, on l'a d'ailleurs réintroduite, notamment aux Etats-Unis où la Commission des Examens d'Admission dans l'Enseignement Supérieur a renoncé au « Test objectif de composition » et en Angleterre, aux deux épreuves-clés de sélection : le 11 + et le 16 + Examination.

Toutefois, si la rédaction continue à jouer un rôle important dans les examens, la façon encore trop répandue de noter doit être modifiée. Nous faisons nôtre la proposition de R. Braddock :

« Si l'on doit évaluer un grand nombre d'étudiants pour décider de leur réussite ou de leur échec, il importe de permettre à ceux qui échouent de faire un second travail (...). Si l'on entend attribuer des grades, trois travaux au moins sont nécessaires ; on retiendra les deux meilleurs dont on fera la moyenne (...). Les travaux doivent être notés par au moins deux correcteurs, utilisant un système de notation, bien compris et accepté, auquel on les a bien entraînés. »¹

III. La notation objective.

La notation objective ne fait pas intervenir l'avis personnel des correcteurs. L'exemple le plus simple est fourni par une question comportant dix multiplications à raison d'un point par résultat exact. De même, la notation d'un test composé d'*items* à choix multiple s'opère par simple comptage du nombre de choix corrects.

De façon plus générale, on appelle notation objective « l'assignation de valeurs numériques à des échantillons comportementaux suffisamment limités, définis et contrôlés pour permettre un accord général parmi les juges ou notateurs »².

Comme nous l'avons remarqué déjà, l'objectivité de la notation ne garantit en rien l'objectivité de l'examen dans son ensemble. Le choix des questions reste, en dernière analyse, toujours subjectif et l'on imagine aisément un examen dont aucune question n'échantillonnerait réellement la capacité que l'on prétend évaluer. Tout objective que puisse être la correction, l'épreuve n'aurait, dans pareil cas, aucune validité. Les exemples de semblable mésaventure fourmillent.

IV. L'étalonnage ou mesure de la position relative.

L'interprétation de toute note soulève un problème de comparabilité.

Etalonner consiste à définir des *normes*, c'est-à-dire des distributions statistiques permettant de situer les performances individuelles.

1 R. BRADDOCK et al., o.c., p. 45.

2 J.C. FLANAGAN, Units, Scores and Norms, in E. LINDQUIST, Ed., *Educational Measurement*, Washington, A.C.E., 1961, 4^e éd.

Selon le but poursuivi, le groupe restreint ou pour comprendre jusqu'à la limite de validité se limite toujours à utiliser, il importe de savoir

Longtemps, le pourcentage norme scolaire. Pour étalonner parfois aussi d'exprimer le pour cent. On sait toutes les

Aujourd'hui, la différence de situation sur une échelle de mesure¹ joue un rôle déterminant dans la constitution des échelles « universelles » de notations sur lesquelles nous obtenons des scores bruts, obtenus et comparables.

Les systèmes d'étalonnage accessibles sont les suivants :

1. Le centilage ;
2. Les notes standards ;
3. L'échelle normalisée ;
4. L'échelle normalisée.

Nous les étudions successivement.

A. Le centilage.

1) Définition.

Le centile, ou rang occurrentiel, jouit encore d'une certaine importance extérieure avec le pourcentage fondamental. Le pourcentage total des points attribués à un centile indique combien d'étudiants ont donné si la classe comptait 100 élèves, valant au 90^e centile est considéré.

Le médian correspond au 50^e centile.

1 Voir nature et faiblesse des échelles de mesure à considérer une échelle à cent degrés.

2 Rappelons que le médian est la valeur qui sépare en deux parties égales : 1 — 2 — 3 — 4

inue à jouer un rôle important
e trop répandue de noter doit
a proposition de R. Braddock :

and nombre d'étudiants pour
ur échec, il importe de per-
aire un second travail (...). Si
trois travaux au moins sont
ux meilleurs dont on fera la
nt être notés par au moins
ème de notation, bien compris
traînés. »¹

objective.

pas intervenir l'avis personnel
us simple est fourni par une
ations à raison d'un point par
on d'un test composé d'*items*
ple comptage du nombre de

n appelle notation objective
ques à des échantillons com-
és, définis et contrôlés pour
mi les juges ou notateurs »².

qué déjà, l'objectivité de la
ectivité de l'examen dans son
s reste, en dernière analyse,
e aisément un examen dont
ait réellement la capacité que
ive que puisse être la correc-
reil cas, aucune validité. Les
ture fourmillent.

de la position relative.

ote soulève un problème de

des *normes*, c'est-à-dire des
ant de situer les performances

Selon le but poursuivi, les normes sont établies pour un groupe restreint ou pour de plus vastes ensembles pouvant comprendre jusqu'à la population d'un pays. Pratiquement, leur validité se limite toujours à un ensemble déterminé et pour les utiliser, il importe de savoir lequel.

Longtemps, le pourcentage a été l'alpha et l'oméga de la norme scolaire. Pour étalonner certains tests, on se contentait parfois aussi d'exprimer le nombre de réponses correctes en pour cent. On sait toutes les faiblesses de pareil système.

Aujourd'hui, la différence entre des individus et non la situation sur une échelle du savoir considérée comme absolue¹ joue un rôle déterminant. La moyenne ou le médian² constituent les points de repère à partir desquels on construit des échelles « universelles », qui, moyennant certaines conditions sur lesquelles nous reviendrons, permettent d'exprimer des scores bruts, obtenus à différentes épreuves, en unités comparables.

Les systèmes d'étalonnage les plus connus et les plus accessibles sont les suivants :

1. Le centilage ;
2. Les notes standards ou notes Z ;
3. L'échelle normalisée à 5 classes ;
4. L'échelle normalisée à 9 classes (Stanines).

Nous les étudions successivement.

A. Le centilage.

1) Définition.

Le centile, ou rang occupé sur une échelle à cent degrés, jouit encore d'une certaine faveur à cause de sa ressemblance extérieure avec le pourcentage, dont il diffère pourtant de façon fondamentale. Le pourcentage indique quelle proportion du total des points attribués à un examen l'élève a obtenu. Le centile indique combien d'élèves se classeraient après un élève donné si la classe comptait cent élèves. Donc un score équivalant au 90^e centile est supérieur à 90 % de la population considérée.

Le médian correspond au 50^e centile.

¹ Voir nature et faiblesse des échelles d'évaluation. Calculer le pourcentage équivaut à considérer une échelle à cent degrés.

² Rappelons que le médian est le point qui divise une série de notes ordonnées en deux parties égales : 1 — 2 — 3 — 4 — 5 ou 1 — 2 — 3 . 4 — 5 — 6.

2) Calcul.

En principe, on ne calcule pas les centiles à partir de moins de cent notes. L'exemple suivant utilise les cent notes qui nous ont déjà servi lors de l'étude de la courbe de Gauss.

Démarche générale :

- Dresser le tableau des classes (en commençant par la classe inférieure) ;
- Indiquer les fréquences (f) ;
- Calculer les fréquences cumulées (f_c) : chaque fréquence est additionnée à celle qui la précède ;
- Appliquer la formule :

$$C_x = l + \frac{(N.C)/100 - f_c}{f_s} \times i.$$

C_x = valeur de la note (x) en centile.

$(N.C)/100$ = le rang de la note correspondant à un centile quelconque.

Exemple : Le 50^e centile correspond à la note médiane, soit $N/2$. Pour les autres, il s'agit d'une simple règle de trois. Le 75^e centile sera donc : $(100 \times 75) : 100 = 75$ (les chiffres sont les mêmes parce que, dans notre exemple, il y a précisément 100 notes).

l = la moyenne entre la limite supérieure de la classe où se trouve le centile cherché et la limite inférieure de la classe qui la précède.

f_c = fréquence cumulée de la classe qui précède celle où se trouve le centile cherché.

f_s = la fréquence simple de l'intervalle où se trouve le centile cherché.

i = intervalle de classe.

Exemple :

Classes	
	24-28
	29-33
	34-38
	39-43
	44-48
	49-53
	54-58
	59-63
	64-68
	69-73
	74-78

Calcul du 50 ^e centile	
$N.C$	100×50
100	100
l	$\frac{49+48}{2} = 48,5$
f_c	44
f_s	22
C_x	$48,5 + \frac{50-44}{22} \times 5 = 49,9$

N.B. Décilage : le 10^e centile = le Ecart semi-interquartile : (75^e

3) Critique.

L'échelle en centiles n'est pas une vaste échelle d'évaluation à rang occupé, mais non sur une distribution est normale, ce qui l'on se rapproche du médian (normalité parfaite).

Le tableau suivant montre

	-2σ	-1σ
Centiles	2 ^e	16 ^e

Exemple :

Classes	Fréquences (f)	Fréquences cumulées (f _c)
24-28	1	1
29-33	2	3
34-38	8	11
39-43	10	21
44-48	23	44
49-53	22	66
54-58	18	84
59-63	9	93
64-68	5	98
69-73	1	99
74-78	1	100

	Calcul du 50 ^e centile	Calcul du 16 ^e centile	Calcul du 84 ^e centile
N.C	$\frac{100 \times 50}{100}$	$\frac{100 \times 16}{100}$	$\frac{100 \times 84}{100}$
100	100	100	100
<i>l</i>	$\frac{49+48}{2} = 48,5$	$\frac{39+38}{3} = 38,5$	$\frac{54+53}{2} = 53,5$
<i>f_c</i>	44	11	66
<i>f_i</i>	22	10	18
<i>C_e</i>	$48,5 + \frac{50-44}{22} \times 5 = 49,9$	$38,5 + \frac{16-11}{10} \times 5 = 41$	$53,5 + \frac{84-66}{18} \times 5 = 58,5$

N.B. Décilage : le 10^e centile = le 1^{er} décile, etc.
 Ecart semi-interquartile : (75^e C — 25^e C)/2.

3) Critique.

L'échelle en centiles n'est, nous l'avons déjà dit, qu'une vaste échelle d'évaluation à 100 degrés. Elle informe sur le rang occupé, mais non sur la distance entre rangs. Or, si la distribution est normale, cette distance rétrécit à mesure que l'on se rapproche du médian (égal à la moyenne dans le cas de normalité parfaite).

Le tableau suivant montre clairement ce phénomène.

	- 2 σ	- 1 σ	0	+ 1 σ	+ 2 σ
Centiles	2 ^e	16 ^e	50 ^e	84 ^e	98 ^e

On constate que 68 centiles sont agglomérés autour de la moyenne, ce qui donne un classement trop grossier pour la majorité du groupe. On risque, par exemple, de considérer qu'une différence importante sépare le 75^e du 25^e centile. L'ancienne norme du pourcentage subsiste tenacement dans les esprits et incline à penser que le 75^e C vaut trois fois le 25^e. Le tableau ci-dessus nous montre qu'il s'agit, en réalité, de deux résultats relativement proches et moyens.

En fait, si les normes en centiles restent utilisées, notamment pour la sélection dans certaines universités américaines, c'est parce que celles-ci ne recrutent que dans la bande des dix ou vingt centiles supérieurs, zone où la discrimination est satisfaisante.

B. Les notes standard ou notes Z.

Un élève a obtenu les notes brutes suivantes¹. Calcul : 22 sur 25 ; Lecture : 72 sur 100 ; Sciences : 26 sur 50.

L'examen de ces notes n'apprend pas grand-chose. Si le professeur est sévère en lecture et beaucoup moins en calcul, le 72/100 en lecture est peut-être plus méritoire que le 22/25 en calcul. De plus, des notes brutes ne permettent pas la comparaison avec d'autres élèves de même niveau pédagogique, par exemple.

Pour rendre la comparaison possible, on exprime ces notes en fonction des écarts types, ce qui permet de les situer sur une même courbe (notes étalonnées ou scores standard).

Supposons que l'on obtienne les résultats suivants :

	Moyenne	Ecart type
Calcul	15	2,5
Lecture	50	10
Sciences	29	5

Calcul des scores standard : $z = (x - M) / \sigma$.

Exemple :

$$(22 - 15) : 2,5 = 2,8 \sigma.$$

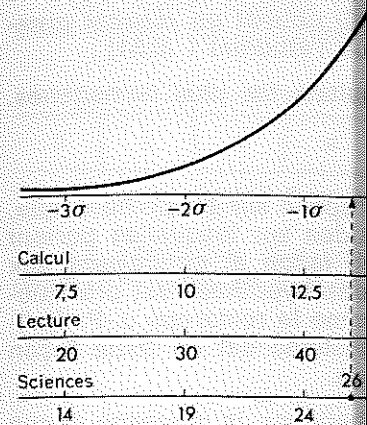
Connaissant z, on peut retrouver la note brute :

$$x = M + z \sigma$$

$$x = 15 + (2,8 \times 2,5) = 22.$$

¹ Adapté d'après R. THOMAS, o.c.

Représentation graphique



Application.

Les étudiants I et II ont

Comparer les résultats

TEST	Moyenne M	Ecart type σ	Score
			I
Anglais	155,7	26,4	195
Lecture	33,7	8,2	20
Information	54,5	9,2	39
Aptitudes scolaires	87,1	25,8	139
Attention	24,8	6,8	41
Totaux Moyennes			434

Si l'on s'en tenait aux scores standard, l'étudiant I serait considéré comme ayant de bons résultats en calcul des pourcentages

¹ Chiffres empruntés à J.-P. GUILLET

sont agglomérés autour de la
 sement trop grossier pour la
 par exemple, de considérer
 are le 75° du 25° centile. L'an-
 subsiste tenacement dans les
 75° C vaut trois fois le 25°. Le
 qu'il s'agit, en réalité, de
 es et moyens.

ntiles restent utilisées, notam-
 aines universités américaines,
 utent que dans la bande des
 zone où la discrimination est

s brutes suivantes¹. Calcul :
 Sciences : 26 sur 50.

prend pas grand-chose. Si le
 et beaucoup moins en calcul,
 plus méritoire que le 22/25 en
 ne permettent pas la compa-
 ème niveau pédagogique, par

ossible, on exprime ces notes
 qui permet de les situer sur
 es ou scores standard).

les résultats suivants :

note	Ecart type
	2,5
	10
	5

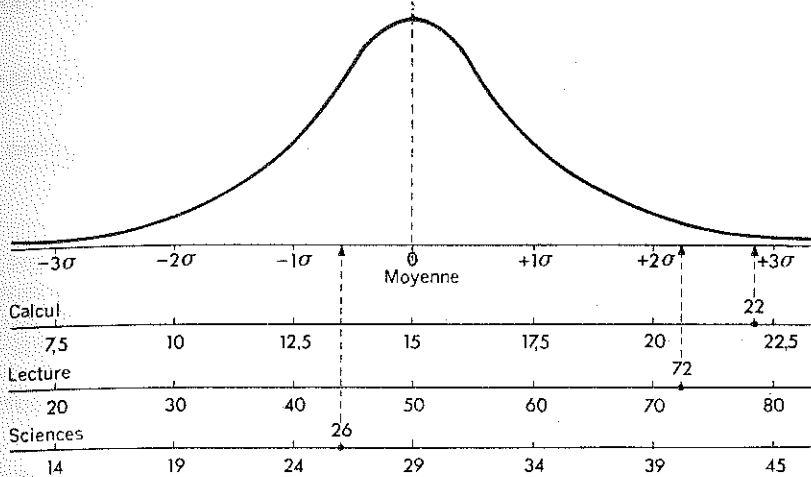
$$= (x - M) / \sigma.$$

$$5 = 2,8 \sigma.$$

er la note brute :

$$\times 2,5) = 22.$$

Représentation graphique :



Application.

Les étudiants I et II ont passé une série de tests.

Comparer les résultats¹.

TEST	Moyen- ne M	Ecart type σ	X		x		Scores	
			Scores bruts		Déviations (M-X)		standards z	
			I	II	I	II	I	II
Anglais	155,7	26,4	195	162	+ 39,3	+ 6,3	+ 1,49	+ 0,24
Lecture	33,7	8,2	20	54	- 13,7	+ 20,3	- 1,67	+ 2,48
Information	54,5	9,2	39	72	- 15,5	+ 17,5	- 1,67	+ 1,88
Aptitudes scolaires	87,1	25,8	139	84	+ 51,9	- 3,1	+ 2,01	- 0,12
Attention	24,8	6,8	41	25	+ 16,2	+ 0,2	+ 2,38	+ 0,03
Totaux Moyennes			434	397			+ 2,54 + 0,51	+ 4,51 + 0,90

Si l'on s'en tenait aux simples totaux (scores bruts I et II), l'étudiant I serait considéré comme supérieur à l'étudiant II. Le calcul des pourcentages confirmerait cette impression.

¹ Chiffres empruntés à J.-P. GUILFORD, *Fundamental Statistics*, o.c., p. 513.

s Z, on ramène tous les résultats à la même unité de déviation par

0,90) obtient un résultat presque l'équivalent de l'étudiant I (0,51) ;

résultats les plus homogènes.

se repose sur l'hypothèse que, les attitudes sur lesquelles portent les notes de la même façon et conduisent aux mêmes résultats. Cette hypothèse est pratique. Comme le remarque J.-P. Guilford, le système est de toute façon valable à tout âge.

l'échelle sur la moyenne donne une notation fautive. En raison des erreurs de notation, le pur hasard fera passer un élève de sujets d'un côté à l'autre.

sur moyenne et un écart type

me peu parlant et préfèrent les quotients intellectuels où la moyenne est de 145 considérée comme la norme.

Cent est donc pris comme la moyenne et divisé par 15.

50 et 100).

ontent un écart type.

moyenne.

due illimitée.

$-1,5\sigma$ $+0,5\sigma$

	4	5
--	---	---

> <-----> <----->
24% 7%

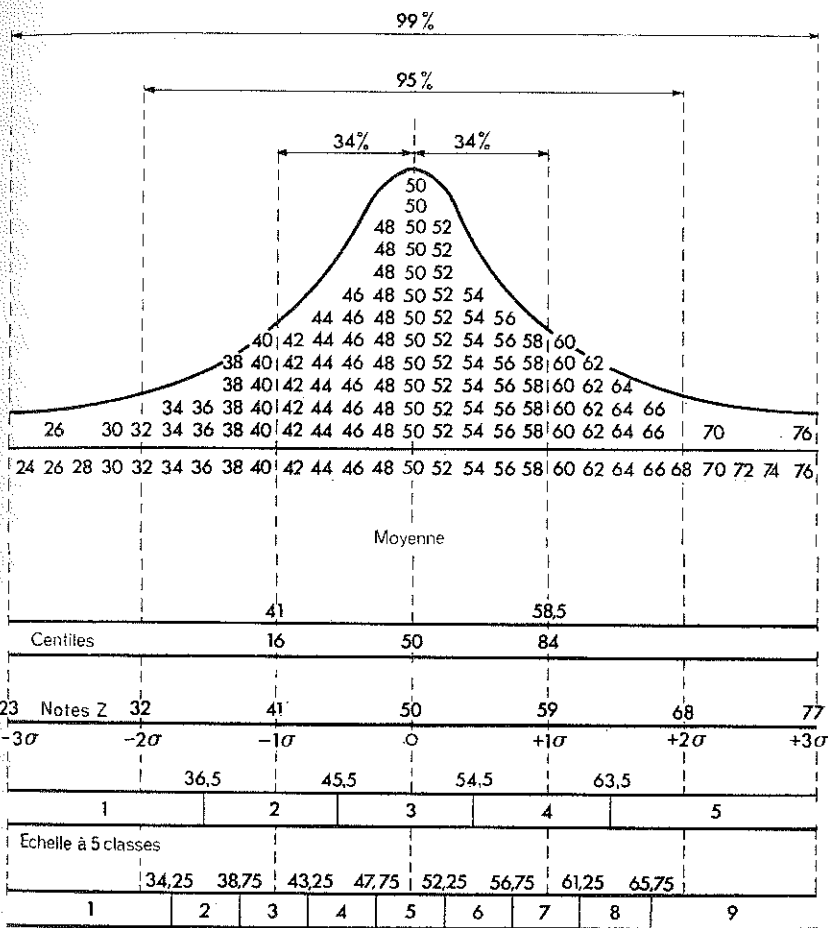
D. L'échelle à 9 classes (Stanines) (N>100).

Les 1^{re} et 9^e classes sont d'étendue illimitée.

Les autres classes comportent un demi-écart type, la 5^e étant centrée sur la moyenne.

	$-1,75\sigma$	$-1,25\sigma$	$-0,75\sigma$	$-0,25\sigma$	M	$+0,25\sigma$	$+0,75\sigma$	$+1,25\sigma$	$+1,75\sigma$
	1	2	3	4	5	6	7	8	9
	>	<----->	<----->	<----->	<----->	<----->	<----->	<----->	<----->
	4%	7%	12%	17%	20%	17%	12%	7%	4%

Le graphique ci-dessous permet une comparaison aisée des quatre types de normes.



Echelle à 9 classes (Stanines)

CHAPITRE 4.

CONTROLE DE LA FIDELITE DE L'EXAMEN.

Beaucoup de mesures que nous avons envisagées jusqu'à présent tendent à assurer une meilleure justice scolaire.

Le contrôle de la fidélité relève de la même préoccupation : idéalement, un même examen (écrit ou oral), passé plusieurs fois, sans que l'élève ait le temps d'apprendre des choses nouvelles, devrait toujours conduire au même résultat. On souhaite donc qu'il soit aussi *fidèle* qu'un mètre qui, bien utilisé, mesure toujours la même longueur, à une erreur minime près.

Evidemment, faire subir la même épreuve à diverses reprises ne permet pas un bon contrôle de la fidélité. Le premier passage provoque une réflexion génératrice d'apprentissage. Par la suite, l'élève continue à penser au problème, vérifie l'exactitude de ses réponses, prend des informations complémentaires, etc. Bref, il faut recourir à un autre moyen ou, plus exactement, à un ensemble de mesures favorables à la fidélité.

En supposant que soit acquise la fidélité de la correction il reste à garantir la fidélité des réponses.

1. Eviter toute ambiguïté dans les questions.

Une question qui manque de clarté au point de se prêter à plusieurs interprétations différentes prive, d'avance, l'examen de sa fidélité. Rien ne permet, en effet, de prévoir avec quelque certitude lequel des sens aurait été donné par un même individu ou par un même groupe, à diverses occasions.

Nous avons envisagé, page 74, un certain nombre de moyens d'assurer la clarté des questions.

2. Des questions en nombre

Si l'on ne pose qu'un...
laisse de vastes zones de...
que l'élève n'ait pas compris...
des parties du cours en spé...
férence accusée du professe...
questions, on enregistrera u...

Le seul moyen d'éviter...
toute la matière. En d'autres...
tillonner l'ensemble.

En principe, plus on a...
augmente.

3. Un contrôle mathématique

Pour des examens impo...
de garanties encore.

a) La méthode pairs-impairs

Si le nombre de questi...
suivante simule une situatio...
tiquement en même temps.

a) Les questions étant nu...
groupes : paires et impair...

b) On dresse deux tablea...
calcule leur corrélation (r...
quement la relation qui e...

c) Les deux examens artifi...
courts que l'original. En...
paragraphe 2, ils sont de...
sous-estime donc la situa...

Une formule simple per...
saire : l'indice de fidélité e...

Plus le résultat sera p...
élevée.

On exige généralement a...

¹ La méthode du calcul de la cor...
statistique élémentaires.

2. Des questions en nombre suffisant.

Si l'on ne pose qu'un très petit nombre de questions, on laisse de vastes zones de matières inexplorées. Or, il se peut que l'élève n'ait pas compris toutes les matières ou ait négligé des parties du cours en spéculant sur la chance ou sur la préférence accusée du professeur. Selon l'endroit où tomberont les questions, on enregistrera un brillant résultat ou un échec.

Le seul moyen d'éviter pareille aventure est de couvrir toute la matière. En d'autres termes, l'examen doit *bien échantillonner* l'ensemble.

En principe, plus on ajoute de questions, plus la fidélité augmente.

3. Un contrôle mathématique.

Pour des examens importants, on souhaitera acquérir plus de garanties encore.

a) La méthode pairs-impairs.

Si le nombre de questions est assez élevé, la démarche suivante simule une situation où un même élève subirait, pratiquement en même temps, deux fois le même examen.

a) Les questions étant numérotées, on les divise en deux groupes : paires et impaires.

b) On dresse deux tableaux parallèles des résultats et l'on calcule leur corrélation (r) ; autrement dit, on évalue numériquement la relation qui existe entre les deux groupes¹.

c) Les deux examens artificiellement créés sont de moitié plus courts que l'original. En vertu de ce que nous avons vu au paragraphe 2, ils sont donc moins fidèles et la corrélation sous-estime donc la situation réelle.

Une formule simple permet d'opérer la correction nécessaire : l'indice de fidélité est finalement obtenu par $\frac{2r}{1+r}$.

Plus le résultat sera proche de 1, plus la fidélité sera élevée.

On exige généralement au moins 0,80.

¹ La méthode du calcul de la corrélation est expliquée dans tous les manuels de statistique élémentaires.

Quand on prépare l'examen par un prétest, une fidélité insuffisante est, notamment, combattue par une augmentation du nombre de questions.

Si le contrôle de fidélité se fait après l'examen effectif, une fidélité trop faible incitera les maîtres à un supplément de prudence dans l'interprétation des résultats et à plus de soin dans la construction des examens futurs.

b) *Deux formes parallèles.*

Le contrôle mathématique est le même que dans la méthode *pairs-impairs*, mais ici les examinateurs préparent deux séries complètes de questions, unanimement jugées équivalentes, à partir de la même grille d'objectifs.

Si la population est assez nombreuse, une moitié, choisie au hasard (par exemple, selon l'ordre alphabétique des noms) reçoit la première forme et l'autre, la seconde. Pour des populations peu nombreuses, on invite les élèves à répondre aux deux séries de questions après quelques jours. Cette seconde solution est évidemment beaucoup moins rigoureuse.

4. Répétition de la notation.

La fidélité des réponses ne suffit pas. Elle doit être accompagnée de la fidélité de la notation. On contrôle celle-ci soit en invitant un même professeur à noter deux fois les mêmes travaux à un intervalle de quelques jours ou de quelques semaines, soit en comparant les notes accordées par plusieurs correcteurs à un même travail.

CONTROL

La fidélité d'un examen mètre mal construit, qui m indiquerait la même longue néanmoins à une conclusion

En bref, valider un exar tivement ce pour quoi il e est capital : car si la cond déployés pour l'organisatio vains ou, au moins, détour

Or, la validation n'est notre savoir sur la validité recherche a à peine touché problème mérite d'autant professeurs.

Selon la nature rétros on distingue la validité du

I. La validité du contenu.

Elle intéresse principa ser le bilan des acquisition Tel but avait été assigné atteints ?

Dans les examens por tissage, on ne peut évide la matière. Non seulemen points principaux, mais e importance relative. Si l'o questions pourront suscite l'ensemble des acquisition

¹ R. COX, *Reliability and Validity Examinations*, o.c., p. 43.

par un prétest, une fidélité
battue par une augmentation

ait après l'examen effectif, une
maîtres à un supplément de
es résultats et à plus de soin
futurs.

le même que dans la méthode
nateurs préparent deux séries
ement jugées équivalentes, à

ombreuse, une moitié, choisie
ordre alphabétique des noms)
e, la seconde. Pour des popu-
te les élèves à répondre aux
quelques jours. Cette seconde
p moins rigoureuse.

uffit pas. Elle doit être accom-
n. On contrôle celle-ci soit en
pter deux fois les mêmes tra-
ours ou de quelques semaines,
cordées par plusieurs correc-

CHAPITRE 5.

CONTROLE DE LA VALIDITE.

La fidélité d'un examen ne garantit en rien sa validité. Un mètre mal construit, qui mesurerait cinq centimètres de trop, indiquerait la même longueur à chaque mesure, mais conduirait néanmoins à une conclusion fautive.

En bref, valider un examen, c'est prouver qu'il mesure effectivement ce pour quoi il est proposé. Acquérir cette certitude est capital : car si la conclusion est négative, tous les efforts déployés pour l'organisation et le passage des examens sont vains ou, au moins, détournés de leur objet.

Or, la validation n'est pas chose aisée. Faisant le total de notre savoir sur la validité, R. Cox conclut, en 1969, que la recherche a à peine touché aux questions fondamentales¹. Le problème mérite d'autant plus l'attention et la prudence des professeurs.

Selon la nature rétrospective ou prospective de l'examen, on distingue la validité du contenu et la validité prédictive.

I. La validité du contenu.

Elle intéresse principalement les examens destinés à dresser le bilan des acquisitions et donc aussi d'un enseignement. Tel but avait été assigné, tels objectifs choisis. Ont-ils été atteints ?

Dans les examens portant sur de longues étapes d'apprentissage, on ne peut évidemment interroger en détail sur toute la matière. Non seulement, il importe donc de localiser les points principaux, mais encore faut-il tenir compte de leur importance relative. Si l'on satisfait à ces deux critères, les questions pourront susciter des réponses *représentatives* de l'ensemble des acquisitions, de la compétence totale.

¹ R. COX, Reliability and Validity of Examinations, in J. LAUWERYS et D. SCANLON, *Examinations*, o.c., p. 43.

En d'autres mots, la validité du contenu dépend de la qualité et de l'adéquation de l'échantillonnage des connaissances et des capacités.

Nous avons vu que, dans un enseignement bien conçu, les objectifs à atteindre, et donc à contrôler lors de l'examen, ont été définis au début de l'année.

Mais, à ce moment, il s'agissait d'intentions que les circonstances peuvent avoir modifiées. Avant d'essayer de dresser le bilan du travail de ses élèves, le professeur fera donc d'abord le sien. Faute de ce retour sur soi-même, les examens n'ont qu'une validité illusoire ou, pire encore, une validité de façade, destinée à dédouaner le professeur aux yeux de ses supérieurs. Ainsi s'explique que des questions portent sur des matières à peine effleurées, mais figurant au programme officiel, ou sur des capacités intellectuelles élevées que l'on a omis d'installer patiemment, systématiquement pendant l'année. Combien d'échecs scolaires ne sont-ils pas dus aux belles questions « d'intelligence », posées au terme d'un enseignement qui ne l'a pas cultivée ?

Si la définition précise des objectifs permet une validation de contenu beaucoup meilleure que dans le passé, celle-ci n'en reste pas moins encore très limitée dans l'état actuel de nos connaissances. D'abord, parce qu'il est bien difficile d'acquérir la certitude que tel comportement réel est bien représentatif de tel trait que l'on ambitionnait de cultiver. Ensuite, parce que, même si la vision est correcte au départ, les vicissitudes de la notation peuvent toujours trahir les intentions.

Un professeur d'histoire ou de géographie, écrit P. Vernon¹ qui, par un examen écrit, essaie de juger la compréhension profonde de ces disciplines, attribue souvent une bonne partie des points pour la reproduction exacte de faits détaillés, la restitution de ses théories favorites, la longueur des réponses, leur intelligibilité et l'élégance du style dans lequel elles sont exprimées. Dans un examen oral, combien les points ne dépendent-ils pas de l'intelligence sociale, du contrôle de soi-même et, spécialement, de l'aptitude de l'élève à créer, par la parole et l'attitude, une bonne relation avec l'examineur ?

Que faire, en pratique ?

D'abord, exiger que, pour chaque question d'examen qu'il rédige, le professeur indique l'objectif poursuivi sur une fiche spéciale.

¹ P. VERNON, Types of Examination, in J. LAUWERYS et D. SCANLON, *Examinations*, p. 43.

Ensuite, soumettre les questions à plusieurs éducateurs expérimentés (qui tiendra à un groupe de discussion) qui jugeront indépendamment des questions. En cas de désaccord, les questions doivent être rejetées¹.

L'expérience montre qu'un petit nombre d'objectifs est préférable à un grand nombre large. Sinon, les juges ne réussissent pas à valider. Nous l'avons dit plus haut, les validations sont encore frustes.

Pour la même raison, un résultat ne sera obtenu sur les questions qu'avec un doute pas un motif suffisant. La validité d'une matière ou de la branche d'accord doit alors être claire.

On risque, par exemple, de proposer du contrôle systématique assignés aux langues anciennes, mathématique, etc.

De nouveau, nous constatons que la rigueur scientifique des examens n'est pas mise en cause. Le fétichisme des notes est en cause.

II. La validité prédictive.

Elle semble, de loin, la plus importante. Elle n'exige pas de compréhension profonde. Même si l'on ne s'explique pas le succès à tel test et une réussite à tel autre, le fait peut être aisément constaté.

Qu'il s'agisse d'un examen de fin de cycle d'études ou autorisant l'inscription de suivre les individus pendant un certain temps, si le pronostic formulé à priori est confirmé ou non.

On constate avec étonnement que la validation est rarement faite. J.-C. Passolunghi est moins destiné à mesurer que l'élève.

¹ Un coefficient de concordance en Kendall et un seuil d'acceptation choisi d'avance.

du contenu dépend de la qualité
de l'enseignement des connaissances et

de l'enseignement bien conçu, les
examinateurs, lors de l'examen, ont

des intentions que les cir-
constances. Avant d'essayer de dresser
un programme officiel, ou sur des
examens n'ont pas encore, une validité de façade,
aux yeux de ses supérieurs.
portent sur des matières à
programme officiel, ou sur des
que l'on a omis d'installer
pendant l'année. Combien
pas dus aux belles questions
de d'un enseignement qui ne l'a

des objectifs permet une validation
que dans le passé, celle-ci n'en
est dans l'état actuel de nos
qu'il est bien difficile d'acquérir
et réel est bien représentatif de
à cultiver. Ensuite, parce que,
au départ, les vicissitudes de la
des intentions.

de géographie, écrit P. Vernon¹
de juger la compréhension
souvent une bonne partie
exacte de faits détaillés, la
des, la longueur des réponses,
le style dans lequel elles sont
combien les points ne dépen-
dent, du contrôle de soi-même
l'élève à créer, par la parole
avec l'examineur ?

chaque question d'examen qu'il
s'agit de poursuivre sur une fiche

LAUWERYS et D. SCANLON, *Examinations*,

Ensuite, soumettre les questions et le schéma de notation
à plusieurs éducateurs expérimentés (dont un au moins appar-
tiendra à un groupe de disciplines différent du domaine con-
cerné) qui jugeront indépendamment de la validité du contenu
des questions. En cas de désaccord grave, la question doit
être rejetée¹.

L'expérience montre qu'il faut actuellement s'en tenir à
un petit nombre d'objectifs généraux conçus de façon assez
large. Sinon, les juges ne réussissent plus à se mettre d'accord.
Nous l'avons dit plus haut, nos connaissances en matière de
validation sont encore frustes.

Pour la même raison, dans certains cas, aucun accord
ne sera obtenu sur les questions fondamentales. Ce n'est sans
doute pas un motif suffisant pour cesser l'enseignement de la
matière ou de la branche incriminée, mais l'impossibilité
d'accord doit alors être clairement indiquée.

On risque, par exemple, de rencontrer pareille difficulté à
propos du contrôle systématique de la réalisation des objectifs
assignés aux langues anciennes, à certaines matières de la
mathématique, etc.

De nouveau, nous constatons qu'en tentant d'accroître la
rigueur scientifique des examens, des enseignements seront
mis en cause. Le fétichisme pédagogique ne disparaîtra qu'à
ce prix.

II. La validité prédictive.

Elle semble, de loin, la plus simple à contrôler, car elle
n'exige pas de compréhension profonde des phénomènes.
Même si l'on ne s'explique pas le rapport existant entre le
succès à tel test et une réussite brillante dans un domaine déter-
miné, le fait peut être aisément observé.

Qu'il s'agisse d'un examen donnant accès à un nouveau
cycle d'études ou autorisant l'exercice d'une profession, il suffit
de suivre les individus pendant un certain temps pour savoir
si le pronostic formulé à partir des notes attribuées se vérifie
ou non.

On constate avec étonnement que semblable vérification
est rarement faite. J.-C. Passeron y voit le signe que l'examen
est moins destiné à mesurer objectivement la capacité qu'à

¹ Un coefficient de concordance entre les avis peut être calculé par la formule de
Kendall et un seuil d'acceptation choisi de commun accord.

servir les besoins *sociaux* des classes privilégiées¹. Cette affirmation mérite certes des réserves, mais nous avons vu, dans la première partie, qu'elle n'est pas entièrement dénuée de fondement.

Pour n'être pas un leurre, la vérification systématique de la validité prédictive exige le contrôle rigoureux des variables, faute de quoi les variables cachées fausseront les conclusions. Par exemple, on tirait encore récemment argument en faveur du latin en faisant la statistique du nombre d'étudiants brillants qui avaient fait des études secondaires classiques. Or, cette observation ne prouve rien, sinon que, traditionnellement dans nos pays, les élèves les plus intelligents sont orientés dès douze ans vers ce type d'études.

En guise de synthèse, nous empruntons la récapitulation suivante à l'*Examination Bulletin* n° 3².

FACTEURS DE VALIDITE D'UN EXAMEN

1. Identification adéquate des objectifs dont la réalisation doit être vérifiée par l'examen écrit, l'évaluation du travail de l'année, l'épreuve pratique, etc.
2. Parmi toute la gamme des objectifs ainsi identifiés, sélection de ceux sur lesquels la vérification se concentrera.
3. Evaluation efficace de l'adéquation du contenu et de la structure de l'examen au but poursuivi.
4. Relation claire entre chaque question et les objectifs de l'enseignement.
5. Elaboration d'un schéma de notation et rédaction de directives aux correcteurs en fonction des objectifs. Un niveau suffisant de fidélité doit aussi être assuré.
6. Bonne connaissance des capacités des candidats.
7. Disposition des notateurs à tenir compte des jugements indépendants du leur et des données objectives qui leur seraient fournies.
8. Comparaison avec des épreuves antérieures dont la validité a été prouvée.

1 J.-C. PASSERON, *Sociologie des examens*, o.c., p. 7.

2 Londres, H.M.S.O., 1964, pp. 19-20.

les classes privilégiées¹. Cette réserve, mais nous avons vu, n'est pas entièrement dénuée de

la vérification systématique de la maîtrise rigoureuse des variables, les fautes fausseront les conclusions. Or, récemment argument en faveur du nombre d'étudiants brillants des secondaires classiques. Or, en, sinon que, traditionnellement les plus intelligents sont orientés dès

us empruntons la récapitulation in n° 3².

DITE D'UN EXAMEN

objectifs dont la réalisation doit être écrite, l'évaluation du travail de

objectifs ainsi identifiés, sélection de la question se concentrera.

évaluation du contenu et de la question poursuivie.

la question et les objectifs de

notation et rédaction de direction des objectifs. Un niveau doit être assuré.

capacités des candidats.

tenir compte des jugements et des données objectives qui leur

études antérieures dont la validité

¹, o.c., p. 7.

QUATRIEME PARTIE

LES PROCEDURES DE MODERATION

POSITION D

1. Définition.

La *modération*, en docimologie, consiste à tempérer les excès de sévérité de certains examinateurs.

Au sens large, elle désigne les mesures prises pour rendre comparables les résultats des examens internes et donc pour unifier les notes des différentes classes de maternelles, d'un groupe d'écoles, dans des régions ou d'un pays.

Les recherches les plus récentes sont certainement dues aux docimologues bretons. En Bretagne, les programmes et les méthodes des écoles primaires ou secondaires peuvent différer considérablement. Pendant très longtemps, des examens étaient organisés en dehors des écoles, par exemple pour sanctionner les études aux collèges. On connaît le fameux 11 + Examen de fin de l'enseignement général classique (Secondary Education), avant l'entrée à l'université. Ce dernier examen, par exemple, est sanctionné par un document signalant le niveau de l'élève par rapport à des normes nationales.

Pour admettre un élève à l'université, les universités annoncent leurs exigences. Les notes d'excellence (A) sont demandées en biologie et en langue maternelle. Les notes d'excellence sont aussi acceptées, etc.

CHAPITRE 1.

POSITION DU PROBLEME.

1. Définition.

La *modération*, en docimologie, a d'abord eu pour objet de tempérer les excès de sévérité ou de générosité chez certains examinateurs.

Au sens large, elle désigne aujourd'hui l'ensemble des mesures prises pour rendre comparables les notes d'examens internes et donc pour unifier leur signification au niveau des différentes classes de même type dans une école, dans un groupe d'écoles, dans des établissements similaires d'une région ou d'un pays.

Les recherches les plus fines sur la modération sont certainement dues aux docimologistes britanniques. Dans un système complètement décentralisé comme celui de la Grande-Bretagne, les programmes et les méthodes des écoles primaires ou secondaires peuvent différer considérablement. Aussi, pendant très longtemps, des examens externes, c'est-à-dire organisés en dehors des écoles, par des commissions spéciales, ont dû sanctionner les études aux moments cruciaux de la scolarité. On connaît le fameux 11 + *Examination*, examen d'entrée dans le secondaire, décidant, en particulier, de l'admission dans l'enseignement général classique, et le C.S.E. (*Certificate of Secondary Education*), avant l'entrée dans l'enseignement supérieur. Ce dernier examen, par exemple, pourvoit chaque élève d'un document signalant le niveau dans chacune des branches, par rapport à des normes nationales.

Pour admettre un élève dans une section déterminée, les universités annoncent leurs exigences particulières : ici, des notes d'excellence (A) sont demandées en mathématiques, en biologie et en langue maternelle ; là, les notes « très bien » (B) sont aussi acceptées, etc.

Ce système d'examens externes est critiqué, parfois très sévèrement, parce qu'il limite gravement la liberté pédagogique des maîtres et des autorités locales.

Aussi, depuis quelques années, on tente de les remplacer par des examens internes dont les résultats sont rendus comparables par des procédures de modération locale, régionale, puis nationale, expérimentalement mises au point.

Dans des pays comme la Belgique, où un programme d'études unique est imposé par une autorité centrale, il est beaucoup plus facile, si on le veut, de modérer les examens. Avant de voir comment, il n'est peut-être pas inutile de rappeler pourquoi une modération est souhaitable.

En gros, la raison est double. La première concerne les individus, élèves et parents, qui ont le droit de connaître le niveau « réel » des performances scolaires, avant de décider de l'orientation ultérieure dans les domaines des études ou de la profession. Qui se croirait qualifié pour les jeux olympiques parce qu'il a gagné une course organisée entre quelques amis ? Un constat défavorable n'est d'ailleurs pas plus une condamnation sans appel qu'un diagnostic de faiblesse ou de maladie. Il faut savoir à temps qu'un problème se pose pour en chercher les causes, puis les remèdes, s'ils existent.

La seconde raison concerne la communauté. De même qu'un consommateur ne peut être trompé sur la marchandise, de même la société ne peut être tenue ou à payer fort cher les échecs d'un étudiant entré à l'université sur la foi d'un certificat invalide, ou à confier une fonction à quelqu'un qui ne la mérite pas. Le jour où, par exemple, les examens de l'enseignement normal seront rigoureusement contrôlés, on peut espérer ne plus voir confier des générations d'enfants à certains instituteurs dont la connaissance de la langue maternelle, notamment de l'orthographe¹, est très insuffisante.

2. Modérer n'est pas caporaliser.

On pourrait craindre que la volonté de rendre les résultats scolaires comparables ne provoque une résurgence des vieilles contraintes. On se souvient de cet inspecteur, de tradition napoléonienne, qui, consultant sa montre, croyait pouvoir dire : « A cet instant, on enseigne telle leçon dans toutes les cinquièmes années primaires de France. » Nous savons les faiblesses d'un enseignement caporalisé.

¹ Nous n'avons nullement le fétichisme de l'orthographe, mais tenons pour évident que celui qui doit l'enseigner doit la connaître.

Mais la nécessité d'un pas qu'au-delà des années cycle d'études poursuit des à tous : acquisition de techniques et de capacités jugées essentielles s'interroger et s'entend

3. Modération volontaire ou

Normaliser tous les examens ne semble ni souhaitable, parce que les élèves sont très différents et s'accommodent en périodes et en années. La modération, même par les examens, reste un problème à être présentées, reste un problème dans son exécution.

La modération doit être décidée avant tout lors de l'attribution de la fin de cycle.

Pour le reste, la décision doit être prise individuellement, soit par les autorités locales.

4. La modération commencée

Nous venons d'y faire allusion. Les modérateurs aient pu trouver dans l'enseignement et sur quel point pour que les examens soient

Des affirmations vagues s'exprimer correctement par une mesure désirée doit être traduite en termes que l'on aura d'ailleurs intérêt à d'évaluation descriptives. Si des questions se posent :

Comment graduer les examens des élèves et du type d'école ?

L'originalité sera-t-elle évaluée ? Comment sera-t-elle identifiée ?

Quelle importance attribuer à la longueur des travaux ?

La longueur des travaux sera-t-elle évaluée ? Exige-t-on une certaine longueur pour laquelle des points seront

externes est critiqué, parfois très gravement la liberté pédagogique cales.

années, on tente de les remplacer et les résultats sont rendus compatibles de modération locale, régionale, ment mises au point.

Belgique, où un programme d'étude e autorité centrale, il est beaucoup modérer les examens. Avant de e pas inutile de rappeler pourquoi ble.

ouble. La première concerne les qui ont le droit de connaître les nces scolaires, avant de décider s les domaines des études ou de qualifié pour les jeux olympiques e organisée entre quelques amis ? d'ailleurs pas plus une condamna- stic de faiblesse ou de maladie. Il blème se pose pour en chercher s'ils existent.

erne la communauté. De même être trompé sur la marchandise, tre tenue ou à payer fort cher les l'université sur la foi d'un certifi- e fonction à quelqu'un qui ne la ample, les examens de l'enseigne- ment contrôlés, on peut espérer rations d'enfants à certains insti- de la langue maternelle, notam- ès insuffisante.

iser.

la volonté de rendre les résultats voque une résurgence des vieilles e cet inspecteur, de tradition napo- montre, croyait pouvoir dire : « A leçon dans toutes les cinquièmes » Nous savons les faiblesses d'un

me de l'orthographe, mais tenons pour évident itre.

Mais la nécessité d'une plasticité pédagogique n'empêche pas qu'au-delà des aménagements circonstanciels, chaque cycle d'études poursuit des objectifs fondamentaux, communs à tous : acquisition de techniques de base ou de connaissances et de capacités jugées essentielles. Maîtres et modérateurs doivent s'interroger et s'entendre sur ces apprentissages cruciaux.

3. Modération volontaire ou imposée ?

Normaliser tous les examens à tous les moments des études ne semble ni souhaitable, ni d'ailleurs possible. Peu souhaitable, parce que les élèves progressent à des rythmes parfois très différents et s'accommodent mal du découpage rigoureux en périodes et en années scolaires. Impossible parce que la modération, même par les méthodes économiques qui vont être présentées, reste un travail lourd, dans sa préparation et dans son exécution.

La modération doit être imposée aux moments décisifs, avant tout lors de l'attribution de diplômes ou de certificats de fin de cycle.

Pour le reste, la décision devrait être laissée soit aux maîtres individuellement, soit aux chefs d'établissement ou aux autorités locales.

4. La modération commence au début de l'année scolaire.

Nous venons d'y faire allusion, il importe que professeurs et modérateurs aient pu trouver un accord sur les objectifs de l'enseignement et sur quelques grands principes de notation pour que les examens soient comparables.

Des affirmations vagues telles que : « L'élève doit pouvoir s'exprimer correctement par écrit » sont sans utilité. La capacité désirée doit être traduite en termes de comportements concrets, que l'on aura d'ailleurs intérêt à incorporer dans les échelles d'évaluation descriptives. S'il s'agit de l'expression écrite, bien des questions se posent :

Comment graduer les exigences en fonction de l'âge des élèves et du type d'école ?

L'originalité sera-t-elle récompensée ? Comment ? Comment sera-t-elle identifiée ?

Quelle importance attribuer à l'orthographe ?

La longueur des travaux de composition sera-t-elle considérée ? Exige-t-on une longueur minimum, en dessous de laquelle des points seront décomptés ? Combien ?

Quelle importance accorder à la richesse du vocabulaire ?

Comment la définir ? Quelles seront les exigences syntaxiques ?

Les mathématiciens peuvent aussi accorder leurs violons. Si, pour nous limiter à un seul exemple, certains professeurs pénalisent gravement les fautes d'opérations, voire les fautes d'orthographe dans la solution de problèmes, alors que leurs collègues estiment que ces aspects sont secondaires, des notes égales risquent de recouvrir des réalités très différentes.

Les programmes officiels peuvent aussi jouer un rôle important en précisant au maximum les buts à atteindre à chaque échelon. L'efficacité de ces données de base est d'autant plus grande que le système scolaire est centralisé.

Inutile de se faire des illusions : les objectifs comportementaux des différentes branches ne se définissent pas en quelques heures de méditation. Des commissions où enseignants, inspecteurs et chercheurs uniront leurs efforts devront travailler longtemps avant d'arriver à un résultat satisfaisant. Beaucoup de questions risquent d'ailleurs de rester provisoirement sans réponse.

La définition des objectifs fera l'objet d'une recherche permanente, non seulement parce qu'ils pourront se préciser en fonction de l'avancement de la psychologie de l'apprentissage, mais aussi parce que les objectifs mêmes évoluent.

Les instructions relatives aux examens se préciseront parallèlement. A côté des grilles d'objectifs généraux et spéciaux, les programmes scolaires de l'avenir détermineront nettement la quantité et les types de questions à poser aux examens et fourniront des indications sur les lignes de force des schémas de notation.

5. Pas de comparabilité sans fidélité élevée.

Il n'est pas concevable de comparer entre eux des résultats d'examens qui, pris isolément, seraient éminemment fluctuants.

Le problème de la fidélité a été discuté dans la partie consacrée à la construction de l'examen. Nous n'y revenons pas.

6. Peut-on se fier aux tests ?

Dans les systèmes de moderner, les tests d'intelligence ou de place importante. En certaines points de repère sur lesquels le en d'autres cas, on leur accorde qu'au travail de l'année et qu'à

Maintenant que nous avons les précautions les bons tests de la confiance accordée à ces inster. Par ailleurs, dès la première avons aussi vu que si les maître simple test verbal, administré en prédiction presque aussi sûre :

	Corrélation 2 ans
Pronostic du maître Test verbal	

Voici encore quelques résultats de ces observations.

A moyen terme, un test d'intelligence est une combinaison de scores à différents niveaux. Les meilleurs prédicteurs du succès scolaire sont les tests de l'enfance. Cette supériorité a été démontrée surtout par des chercheurs anglo-américains.

- 1) Emmet¹ montre qu'un test d'intelligence peut mieux prévoir les résultats, après cinq ans, que le test de l'âge scolaire, le test de l'âge maternel et d'arithmétique élémentaire.
- 2) Emmet et Wilmut² ont, par ailleurs, obtenu un résultat tout aussi convaincant de la validité de la prédiction de l'intelligence, cinq ans après.

¹ W. EMMET, *An Inquiry into the Prediction of School Achievement*, Univ. of London Press, 1942.
² W. EMMET and F. WILMUT, *The Prediction of School Achievement in Specific Subjects*, in *British Journal of Educational Psychology*, 1942.

der à la richesse du vocabulaire ?

uelles seront les exigences syn-

vent aussi accorder leurs violons.

seul exemple, certains professeurs

ites d'opérations, voire les fautes

on de problèmes, alors que leurs

spects sont secondaires, des notes

des réalités très différentes.

ls peuvent aussi jouer un rôle

maximum les buts à atteindre à cha-

ces données de base est d'autant

scolaire est centralisé.

sions : les objectifs comporte-

ne se définissent pas en quelques

missions où enseignants, inspec-

teurs efforts devront travailler long-

résultat satisfaisant. Beaucoup de

de rester provisoirement sans

s fera l'objet d'une recherche per-

ce qu'ils pourront se préciser en

la psychologie de l'apprentissage,

ectifs mêmes évoluent.

aux examens se préciseront paral-

d'objectifs généraux et spéciaux,

l'avenir détermineront nettement

uestions à poser aux examens et

les lignes de force des schémas

fidélité élevée.

comparer entre eux des résultats

, seraient éminemment fluctuants.

ité a été discuté dans la partie

de l'examen. Nous n'y revenons

6. Peut-on se fier aux tests ?

Dans les systèmes de modération que nous allons examiner, les tests d'intelligence ou de connaissances occupent une place importante. En certaines occasions, ils apportent les points de repère sur lesquels les autres résultats sont ajustés ; en d'autres cas, on leur accorde, à eux seuls, autant de valeur qu'au travail de l'année et qu'à l'examen final.

Maintenant que nous avons vu avec quelle rigueur et quelles précautions les bons tests de connaissances sont construits, la confiance accordée à ces instruments ne doit plus nous étonner. Par ailleurs, dès la première partie de ce livre, nous avons aussi vu que si les maîtres jugent bien leurs élèves, un simple test verbal, administré en moins d'une heure, permet une prédiction presque aussi sûre :

	Corrélation avec résultats de l'élève :	
	2 ans après	3 ans après
Pronostic du maître	. 821	. 748
Test verbal	. 796	. 722

Voici encore quelques résultats de recherches confirmant ces observations.

A moyen terme, un test d'intelligence ou, mieux encore, une combinaison de scores à différents tests d'intelligence, sont meilleurs prédicteurs du succès scolaire que les résultats d'exams. Cette supériorité a été démontrée à plusieurs reprises, surtout par des chercheurs anglo-saxons :

- 1) Emmet¹ montre qu'un test d'intelligence verbale permet de mieux prévoir les résultats, après deux ou trois ans d'enseignement secondaire général, que les examens de langue maternelle et d'arithmétique évalués par les maîtres.
- 2) Emmet et Wilmut² ont, par la suite, fait une démonstration tout aussi convaincante de la valeur prédictive du test d'intelligence, cinq ans après.

¹ W. EMMET, *An Inquiry into the Prediction of Secondary School Success*, London, Univ. of London Press, 1942.

² W. EMMET and F. WILMUT, *The Prediction of School Certificates Performance in Specific Subjects*, in *British Journal of Educ. Psychol.*, 22, 1952, 52-62.

- 3) Wrigley¹ a confirmé ces résultats et montré que la prédiction peut être meilleure encore si les résultats à un test d'intelligence et à des tests de connaissances standardisés peuvent être combinés.
- 4) En Belgique, tous les orienteurs connaissent la valeur prédictive des scores verbaux et de raisonnement du Test P.M.A. de Thurstone, le fameux 2 V + R, pour l'enseignement secondaire général.

Nous ne disposons pas de recherches similaires pour l'enseignement technique ou pour des structures d'enseignement peut-être mieux adaptées à la civilisation de l'an 2000. Il est très possible que les scores d'autres types de tests devraient être utilisés.

¹ J. WRIGHLEY, The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools, in *British Journal of Educ. Psychol.*, 25, 1955, 107-116.

QUELQUES SYSTEMES DE M

I. La formule la

Le système suédois de à partir d'un test

Dès la fin de la seconde guerre mondiale, un système de modération, similaire à celui de chaque professeur, dans sa classe.

Toutes les notes des professeurs sont converties sur une échelle de 7 degrés¹.

On suggère que le pourcentage déterminé soit, en gros, le suivant : on doit tenir compte des caractéristiques de tête très forte, ou presque tous

Notes	1
Pourcentage	1

Par ailleurs, une batterie de tests est administrée tant sur les branches principales que sur les autres chaque année par un office centralisé de spécialistes et des professeurs expérimentés. Les résultats sont répartis en 7 classes et les notes des professeurs.

¹ Nous nous référons à l'exposé très intéressant de la *System of Equalising Marks*, in *Educational Research*, 1955, 107-116.

² Distribution normale (voir courbe de Gauss)

ésultats et montré que la prédic-
ncore si les résultats à un test
s de connaissances standardisés

teurs connaissent la valeur pré-
x et de raisonnement du Test
eux 2 V + R, pour l'enseigne-

recherches similaires pour l'en-
des structures d'enseignement
ivilisation de l'an 2000. Il est très
es types de tests devraient être

CHAPITRE 2.

QUELQUES SYSTEMES DE MODERATION DES EXAMENS.

I. La formule la plus libérale :

Le système suédois de modération par branche à partir d'un test de connaissances.

Dès la fin de la seconde guerre mondiale, la Suède a adopté un système de modération, simple et facultatif, applicable par chaque professeur, dans sa classe, tout au long de la scolarité.

Toutes les notes des professeurs sont attribuées selon une échelle de 7 degrés¹.

On suggère que le pourcentage d'élèves recevant une note déterminée soit, en gros, le suivant², le professeur restant libre de tenir compte des caractéristiques de sa classe (par exemple : tête très forte, ou presque tous élèves moyens, etc.) :

Notes	1	2	3	4	5	6	7
Pourcentage	1	6	24	38	24	6	1

Par ailleurs, une batterie de tests de connaissances portant sur les branches principales du programme est construite chaque année par un office central où collaborent des spécialistes et des professeurs expérimentés. Cette batterie est étalonnée sur un échantillon national représentatif ; les résultats sont répartis en 7 classes, dans la proportion conseillée pour les notes des professeurs.

of Intelligence and Attainment Tests as
British Journal of Educ. Psychol., 25, 1955,

¹ Nous nous référons à l'exposé très clair de S. HENRYSSON, The Swedish System of Equalising Marks, in *Educational Research*, VI, 2, Feb. 1964, 156-160.

² Distribution normale (voir courbe de Gauss).

Si le professeur administre le test national dans sa classe, il dispose donc de normes directement comparables aux notes qu'il a attribuées. Il lui est donc facile d'ajuster ces dernières.

Voici un exemple concret proposé par S. Henrysson :

24 élèves			
Notes	Distribution des notes préliminaires attribuées par le professeur	Distribution des notes obtenues au test	Distribution des notes après ajustement
7	—	1	—
6	—	4	4
5	4	4	4
4	4	9	10
3	14	5	4
2	2	1	2
1	—	—	—
Moyenne	3,42	4,33	4,17

Le professeur inscrit d'abord, dans la 2^e colonne, les notes qu'il a attribuées. La moyenne (3,42) fournit une première indication, si la moyenne nationale est connue. (En Suède, elle est de 4 pour la 6^e primaire. Il semble donc que le professeur soit ici trop sévère).

Voici maintenant les résultats obtenus au test par les mêmes élèves :

Score brut	Note correspondante	Nombre d'élèves obtenant cette note
94-100	7	1
84-93	6	4
67-83	5	4
46-66	4	9
30-45	3	5
22-29	2	1
0-21	1	—

Ces résultats sont reportés dans le tableau ci-dessous (colonne 3).

La moyenne est 4,33, ce qui correspond à une sévérité de 4,33.

Sans changer l'ordre des notes, le professeur peut rectifier ses notes (colonne 4).

Le système suédois séduit :

1^o Les maîtres gardent leur esprit critique :

- de faire passer ou non les notes ;
- de tenir compte des résultats ;
- de communiquer les résultats aux autres professeurs, au directeur, etc.

Ainsi, le test devient simple et efficace pour la sélection des praticiens. La tentative est conduite au minimum.

2^o Les constructeurs des tests ont une idée précise de l'ampleur de l'échantillon.

Il est évident que les professeurs ont une préférence particulière aux matières traitées et qu'ils ont ainsi d'un moyen efficace pour sélectionner les innovations.

Remarquons que, depuis 1969, ce système a été expérimenté dans des classes de 6^e primaire (10-12 ans) pour l'orientation.

Parmi les critiques formulées, on retiendra :

- Si un professeur a jugé une sévérité excessive, la note ne réparera pas l'injustice.
- L'épreuve de référence présente des dimensions que les notes ne peuvent avoir.
- La construction de l'épreuve est parfois défectueuse : fidélité trop basse.

1 Voir S. ROLLER, Le problème de la normalisation des notes, Genève, in *Docimologie et évaluation de l'éducation*, avril-septembre 1969, pp. 60-61.

2 W. ANGOFF, Can Usual General Admission Tests, in *A. A. B. Different College Admission Tests*, in A. A. B. Washington, ACE, 1966, pp. 251-264.

F. BACHER, La normalisation des notes.

le test national dans sa classe, etement comparables aux notes c facile d'ajuster ces dernières.

proposé par S. Henrysson :

Élèves	
Distribution des notes obtenues au test	Distribution des notes après ajustement
1	—
4	4
4	4
9	10
5	4
1	2
—	—
4,33	4,17

d, dans la 2^e colonne, les notes (3,42) fournit une première indication est connue. (En Suède, elle est ble donc que le professeur soit

tats obtenus au test par les

pondante	Nombre d'élèves obtenant cette note
	1
	4
	4
	9
	5
	1
	—

Ces résultats sont reportés dans le premier tableau, colonne 3.

La moyenne est 4,33, ce qui confirme la première impression de sévérité.

Sans changer l'ordre du classement initial, le professeur rectifie ses notes (colonne 4).

Le système suédois séduit pour plusieurs raisons :

1^o Les maîtres gardent leur entière liberté :

- a) de faire passer ou non le test (presque tous le font) ;
- b) de tenir compte des résultats ;
- c) de communiquer les résultats des tests aux élèves, aux autres professeurs, au directeur, aux parents.

Ainsi, le test devient simplement un outil mis à la disposition des praticiens. La tentation de « bachotage » est donc réduite au minimum.

2^o Les constructeurs des tests veillent à couvrir une large gamme d'objectifs échantillonnant bien tout le programme.

Il est évident que les professeurs prêtent une attention particulière aux matières traitées dans les tests. On dispose donc ainsi d'un moyen efficace pour sensibiliser le maître à certaines innovations.

Remarquons que, depuis 1965, un système fort semblable est expérimenté dans des classes genevoises de 5^e et de 6^e primaire (10-12 ans) pour l'orthographe et l'arithmétique¹.

Parmi les critiques formulées contre ce système d'ajustement², on retiendra :

1. Si un professeur a jugé un des élèves de sa classe avec une sévérité excessive, la référence à l'épreuve nationale ne réparera pas l'injustice.
2. L'épreuve de référence peut ne pas mesurer les mêmes dimensions que les notes.
3. La construction de l'épreuve de référence peut aussi être défectueuse : fidélité trop basse,...

¹ Voir S. ROLLER, Le problème de l'attribution des notes scolaires. Essai de solution, Genève, in *Docimologie et éducation*, numéro spécial de la revue *Les sciences de l'éducation*, avril-septembre 1969, pp. 66 sq.

² W. ANGOFF, Can Usual General Purpose Equivalency Tables be Prepared for Different College Admission Tests, in A. ANASTASI, Ed., *Testing Problems in Perspective*, Washington, ACE, 1966, pp. 251-264.

F. BACHER, *La normalisation des notes*, o.c., p. 63.

II. Système imposé de modération par branche à partir d'un test de connaissances.

A partir des mêmes données de base que celles de la Suède (examens internes et tests de connaissances, pour les branches principales, étalonnés nationalement ou régionalement), le système directif suivant peut donner de bons résultats.

Une commission nationale ou régionale de modération prend un certain nombre d'écoles en charge.

Ces écoles lui envoient les résultats aux épreuves préparées librement par les professeurs et les scores obtenus aux tests de connaissances.

En cas de différence, en plus ou en moins, égale ou supérieure à deux écarts types (par exemple) entre les moyennes aux examens et aux tests, l'école reçoit la visite des modérateurs.

Leur mission n'est pas de dire au directeur comment il doit conduire son école, ni au professeur comment faire son cours, mais bien d'attirer l'attention sur un fait et de tâcher de trouver, en pleine collaboration avec l'école, l'explication et, si possible, le remède.

Si le désaccord subsiste, le droit d'ajuster d'office les notes peut être donné aux autorités régionales, l'école ayant, de son côté, le droit d'interjeter appel.

III. Modération volontaire par appel à une banque d'items.

Dans le cas où les maîtres utilisent des questions sous forme d'items à choix multiple, la démarche suivante, actuellement expérimentée en Grande-Bretagne semble pleine de promesses¹. Toutefois, elle requiert l'existence d'un service de recherche, nécessité dont un éducateur averti ne peut d'ailleurs plus douter aujourd'hui.

1. Les maîtres indiquent sur une grille les objectifs poursuivis.
2. Ils envoient cette grille au service de recherche, en même temps que les items qu'ils ont rédigés et, éventuellement, déjà prétestés localement.
3. Le service examine ces items et, selon les possibilités, met certains d'entre eux à l'épreuve ; leur difficulté et leur pouvoir discriminatif sont calculés.

¹ Voir D. PIDGEON et A. YATES, *o.c*

4. Le service renvoie au maître :

- a) les items examinés ;
- b) des items complémentaires d'efficacité sont connus ville, canton, pays, ... Ces items fourniront un duquel le résultat enregistré est ajusté.

Il est intéressant de noter qu'au Royaume-Uni, dans la rédaction d'items à choix multiple, les Britanniques comme les Américains ont des items plus généralement formulées, pour les besoins de la recherche, mentalement les grands types de questions, et quelles des barèmes de correction.

L'organisation d'une banque d'items pour les premières années. Par la suite, elle devient telle que le travail s'allège et qu'il y a une grande souplesse et une grande variété.

IV. Un système simple de modération au service de la recherche.

Nous avons déjà observé que les maîtres ne savent pas bien leurs élèves, mais qu'ils tentent de le faire par rapport au niveau global de la classe.

Le système suivant, mis au point par la recherche anglaise pour la Recherche en Éducation, est dans l'enseignement secondaire une méthode de sélection en tenant compte de la performance, notamment par les maîtres, dans les classes.

1. L'instituteur classe ses élèves en fonction de leur performance (dire selon son évaluation globale de la classe dans l'enseignement secondaire, les élèves peuvent être classés ex aequo).
2. Les élèves subissent un test de connaissances.
3. Les scores au test sont classés en fonction de la performance (1) et (3).
4. Les classements (1) et (3) sont comparés.

¹ Voir A. YATES et D. PIDGEON, *Admission*

modération par branche de connaissances.

nées de base que celles de la
ests de connaissances, pour les
és nationalement ou régionale-
ant peut donner de bons résul-

e ou régionale de modération
es en charge.

s résultats aux épreuves prépa-
eurs et les scores obtenus aux

us ou en moins, égale ou supé-
r exemple) entre les moyennes
ble reçoit la visite des modéra-

dire au directeur comment il
professeur comment faire son
ntion sur un fait et de tâcher
tion avec l'école, l'explication

le droit d'ajuster d'office les
és régionales, l'école ayant, de
el.

appel à une banque d'items.

s utilisent des questions sous
a démarche suivante, actuelle-
retagne semble pleine de pro-
t l'existence d'un service de
ducateur averti ne peut d'ail-

grille les objectifs poursuivis.

rvice de recherche, en même
nt rédigés et, éventuellement,

et, selon les possibilités, met
ve ; leur difficulté et leur pou-
s.

4. Le service renvoie au maître :

- a) les *items* examinés ;
 - b) des *items* complémentaires dont les indices de facilité et d'efficacité sont connus pour une population donnée : ville, canton, pays,...
- Ces *items* fourniront un point de comparaison à partir duquel le résultat enregistré pour les autres pourra être ajusté.

Il est intéressant de noter qu'après s'être étroitement cantonnés dans la rédaction d'*items* à réponse fermée, par choix multiple, les Britanniques commencent à proposer des questions semi-ouvertes. Il s'agit, en fait, de questions ouvertes, très soigneusement formulées, pour lesquelles on a identifié expérimentalement les grands types de réponses, en fonction desquelles des barèmes de correction sont proposés.

L'organisation d'une banque d'*items* est lourde pendant les premières années. Par la suite, la provision d'*items* étalonnés devient telle que le travail s'allège et que le système acquiert une grande souplesse et une grande rapidité de fonctionnement.

IV. Un système simple de modération globale au service de la sélection.

Nous avons déjà observé qu'en général, les maîtres jugent bien leurs élèves, mais qu'ils tendent à relativiser leurs jugements par rapport au niveau global de la classe.

Le système suivant, mis au point par la Fondation nationale anglaise pour la Recherche en Education¹ pour l'admission dans l'enseignement secondaire général classique, permet de sélectionner en tenant compte des jugements portés indépendamment par les maîtres, dans leur classe.

1. L'instituteur classe ses élèves par ordre de mérite, c'est-à-dire selon son évaluation globale de la chance de réussite dans l'enseignement secondaire général. Plusieurs élèves peuvent être classés *ex aequo* (Classement 1).
2. Les élèves subissent un test d'intelligence verbale (2).
3. Les scores au test sont classés par ordre décroissant (Classement 3).
4. Les classements (1) et (3) sont placés côte à côte.

¹ Voir A. YATES et D. PIDGEON, *Admission to Grammar Schools*, o.c.

5. Le score d'intelligence qui tombe en face du nom de l'élève est considéré comme score d'évaluation par l'instituteur (Jugement ajusté).

Classement par l'instituteur (1)	Score obtenu par l'élève au test verbal (2)	Classement des scores par ordre décroissant (3)	Jugement ajusté (4)
A	121	132	132
B	120	128	128
C	132	121	121
D	128	120	120
E	100	106	106
F	106	100	100
G	94	100	97
H	82	96	97
I	96	94	97
J	100	86	86
K	78	82	78
L	79	79	78
M	86	78	78
N	73	73	78
O	65	65	65
		Total 290	
		Total 312	

On voit que le classement auquel on aboutit finalement ne modifie en rien l'ordre initialement choisi par l'instituteur. Mais, cette fois, nous disposons d'un moyen de comparaison entre écoles.

Utilisation des jugements ajustés

Exemple : Dans un lycée, latine, 25 élèves provenant de

	Ecole primaire I	Ecole
	132	
130		
	128	
125		
	121	
120	120	
115	115 x 3	
	112	
110		
	106	
105		
	100	
100		

Pour sélectionner 25 élèves Si, à la coupure de 110, des 25 élèves dans la catégorie de l'école élémentaire aurait exigé qu'on

Tracer une ligne de démarcation est une source d'injustice pour le seul n'est-il pas responsable de la sélection de sujets juste au-dessus ou

Le psychométricien est autorisé : il prend une marge de sécurité d'erreur standard qu'une formule simple permet de calculer à l'aide du coefficient de fidélité du test utilisé.

¹ Erreur standard = $\sigma \sqrt{1-r}$
 où σ = écart type des scores.
 r = coefficient de fidélité. Ce coefficient est donné dans le manuel qui accompagne le test.

tombe en face du nom de l'élève
re d'évaluation par l'instituteur

Classement des scores par ordre décroissant (3)		Jugement ajusté (4)
132		132
128		128
121		121
120		120
106		106
100		100
100	Total 290	97
96		97
94		97
86		86
82	Total 312	78
79		78
78		78
73		78
65		65

auquel on aboutit finalement ne
ment choisi par l'instituteur. Mais,
n moyen de comparaison entre

Utilisation des jugements ajustés.

Exemple : Dans un lycée, on souhaite recruter, pour la 6^e
latine, 25 élèves provenant de trois écoles primaires différentes.

Jugements ajustés

	Ecole primaire I	Ecole primaire II	Ecole primaire III
	132		
130	-----		----- 130 x 2
	128		
125	-----	125	----- 126
		124	----- 125
	121		
120	----- 120	----- 121	----- 123
		119	----- 120
		117	
115	----- 115 x 3	----- 115	----- 116
	112		
110	-----	----- 110	----- 114
			----- 113
	106	107	
105	-----	----- 104	
100	----- 100		

Pour sélectionner 25 élèves, on coupe, dans ce cas, à 110.
Si, à la coupure de 110, des *ex aequo* avaient placé plus de
25 élèves dans la catégorie des sélectionnés, la justice la plus
élémentaire aurait exigé qu'on les admette tous.

Tracer une ligne de démarcation comme on vient de le faire
est une source d'injustice pour une autre raison : le hasard
seul n'est-il pas responsable du placement d'un certain nombre
de sujets juste au-dessus ou en dessous de la limite ?

Le psychométricien est armé pour surmonter cette difficul-
té : il prend une marge de sécurité de trois fois l'erreur stan-
dard qu'une formule simple permet de calculer à partir du
coefficient de fidélité du test utilisé ¹.

¹ Erreur standard = $\sigma \sqrt{1-r}$

où σ = écart type des scores.

r = coefficient de fidélité. Ce coefficient est généralement indiqué dans le
manuel qui accompagne le test.

V. En Angleterre, un système de modération complet.

Le système que nous allons décrire maintenant est, lui aussi, dû à la Grande-Bretagne. A notre connaissance, il n'en existe pas de plus complet : il porte sur la préparation de l'examen en collaboration avec les écoles, la notation, l'ajustement des notes et des grades finaux.

Le but poursuivi est de *perfectionner les examens internes*, au point de pouvoir leur confier le rôle joué jusqu'à présent par les grandes épreuves externes.

Pour en arriver à une modération nationale, on procède par paliers : à l'échelon local où un certain nombre d'écoles poursuivant une même finalité se groupent ; à l'échelon régional ensuite, selon une technique qui ne diffère pas fondamentalement de celle appliquée localement ; pour le passage au niveau national, s'ajoute, à la technique de modération proprement dite, un échantillonnage fin dont l'étude technique serait hors de propos ici.

Pour notre pays, nous suggérons que quelques écoles de même esprit commencent par faire une expérience volontaire du système. C'est pourquoi nous concentrons notre présentation sur le processus de base.

La modération inter-écoles ne serait faite qu'aux moments cruciaux de la scolarité et pour quelques branches principales. Rien n'empêche, toutefois, les professeurs d'une même école, enseignant les mêmes cours, d'utiliser spontanément la même méthode de correction.

A. Préliminaires.

Nous avons déjà fait allusion aux accords à prendre sur les objectifs et sur la construction de l'examen. Nous n'y revenons plus.

Chaque école choisit en son sein un professeur qui jouera deux rôles :

- 1° Coordonnateur des examens dans son école ;
- 2° Membre de la commission de modération inter-écoles.

Un portrait idéal du modérateur n'existe pas. Les qualités suivantes paraissent souhaitables¹ :

1. Maturité générale et bonne expérience pédagogique ;

¹ MATHER, o.c., p. 67.

2. Contacts fréquents avec des d'école concernés par l'examen ;
3. Capacité de formuler clairement ;
4. Bonté mais fermeté ;
5. Capacité de discuter sans ;
6. Aptitude à comprendre quelque chose de statistique.

Dans chaque école, les questions avec une grande liberté de mouvement objectifs principaux, le nombre et à respecter selon la convention ;

Au début au moins, il est questions à la commission de l'unité générale.

B. Les professeurs notent leurs

C. Nouvelle correction d'un échantillon de copies par les modérateurs

Trois principes dominent le

- 1° L'intervention dans les examens discrète que possible. Seuls les tantes méritent l'attention.
- 2° Les correcteurs doivent, en apprécier les mêmes qualités de notation des grades finaux.
- 3° Les échantillons à recorriger statistique doit être aussi sim

Dans chaque école et pour préleve, au hasard, 20 copies corr

Soit le cas de 12 lycées¹. Ch à la commission, laquelle reçoit

La première opération vise à modérateurs à trois points de vue

- 1° *Sévérité* : son degré est révélé
- 2° *Discrimination* : notation trop Révélée par la dispersion ou

¹ Exemple emprunté à l'*Examinations Bulletin*

de modération complet.

décrire maintenant est, lui
notre connaissance, il n'en
sur la préparation de l'exa-
mes, la notation, l'ajustement

onner les examens internes,
ôle joué jusqu'à présent par

ation nationale, on procède
un certain nombre d'écoles
groupent ; à l'échelon régio-
ui ne diffère pas fondamen-
ment ; pour le passage au
nique de modération propre-
ont l'étude technique serait

ons que quelques écoles de
e une expérience volontaire
ncentrons notre présentation

serait faite qu'aux moments
quelques branches principales.
esseeurs d'une même école,
iser spontanément la même

aux accords à prendre sur
de l'examen. Nous n'y reve-

in un professeur qui jouera

as son école ;
modération inter-écoles.

ur n'existe pas. Les qualités

érience pédagogique ;

2. Contacts fréquents avec des élèves du niveau et du type d'école concernés par l'examen ;
3. Capacité de formuler clairement ses critères et ses jugements ;
4. Bonté mais fermeté ;
5. Capacité de discuter sans passion, sans créer de tension ;
6. Aptitude à comprendre quelques techniques d'analyse statistique.

Dans chaque école, les questions d'examens sont rédigées avec une grande liberté de mouvement : seuls la grille des objectifs principaux, le nombre et le volume des questions sont à respecter selon la convention prise.

Au début au moins, il est souhaitable de soumettre les questions à la commission de modération qui s'assure de l'unité générale.

B. Les professeurs notent leurs examens.

C. Nouvelle correction d'un échantillon de copies par les modérateurs.

Trois principes dominent le travail :

- 1° L'intervention dans les examens des écoles doit être aussi discrète que possible. Seuls les cas de divergences importantes méritent l'attention.
- 2° Les correcteurs doivent, en gros, être de même sévérité, apprécier les mêmes qualités et être d'accord sur la signification des grades finaux.
- 3° Les échantillons à recorriger doivent être petits, et le travail statistique doit être aussi simple que possible.

Dans chaque école et pour un même type d'examen, on prélève, au hasard, 20 copies corrigées.

Soit le cas de 12 lycées¹. Chacun a délégué un modérateur à la commission, laquelle reçoit donc 12 paquets de 20 copies.

La première opération vise à déterminer l'équivalence des modérateurs à trois points de vue :

- 1° *Sévérité* : son degré est révélé par la moyenne.
- 2° *Discrimination* : notation trop désinvolte ou trop prudente. Révélée par la dispersion ou marge de variation des notes.

¹ Exemple emprunté à l'*Examinations Bulletin* n° 5, Londres, HMSO, 1965.

3° *Conformité* : un même élève est-il classé de la même façon par tous les correcteurs ? Révélée par la corrélation entre deux séries de notes.

Pour vérifier l'accord entre modérateurs, à ces trois points de vue, les douze recorrigent d'abord un même paquet de 20 copies.

Les 20 copies sont réparties, toujours au hasard, en cinq groupes de quatre. Dans chaque groupe, les élèves sont classés par ordre alphabétique.

On fait ensuite les simples opérations suivantes, dans l'ordre où elles figurent dans le tableau :

ETUDE DE L'ACCORD ENTRE LES MODERATEURS

Notes accordées par 12 modérateurs

Nom du candidat	MODERATEURS												Moyenne
	a	b	c	d	e	f	g	h	i	j	k	l	
A	4	4	4	3	4	4	3	1	4	4	4	4	4
B	2	2	2	2	3	2	1	1	2	2	2	2	2
C	6	5	5	5	5	5	5	5	5	5	5	5	5
D	2	1	1	2	2	3	3	2	2	3	2	3	2
Total	14	12	12	12	14	14	12	9	13	14	13	14	13
Marge de variat.	4	4	4	3	3	3	4	4	3	3	3	3	3
E	3	3	3	4	3	5	3	3	3	3	3	3	3
F	1	1	1	1	1	1	1	1	1	1	1	1	1
G	5	6	6	6	5	3	6	6	6	6	6	6	6
H	3	4	4	4	3	4	3	4	4	2	4	4	4
Total	12	14	14	15	12	15	13	14	14	12	14	14	14
Marge de variat.	4	5	5	5	4	4	5	5	5	5	5	5	5
I	2	2	2	2	4	3	2	3	3	3	3	3	3
J	2	1	2	1	3	2	2	2	2	2	2	2	2
K	5	5	5	6	5	5	5	6	5	5	5	4	5
L	1	2	2	1	2	2	2	3	2	2	3	2	2
Total	10	10	11	10	14	12	11	14	12	12	13	11	12
Marge de variat.	4	4	3	5	3	3	3	4	3	3	3	2	3
M	4	2	3	4	3	4	5	5	5	4	5	4	4
N	4	3	4	4	4	5	4	4	4	4	4	4	4
O	5	4	4	3	4	3	4	4	4	5	4	5	4
P	5	6	5	5	5	5	5	5	5	5	5	5	5
Total	18	15	16	16	16	17	18	18	18	18	18	18	17
Marge de variat.	1	4	2	2	2	2	1	1	1	1	1	1	1
Q	1	2	1	1	1	1	1	1	1	1	2	1	1
R	1	1	2	1	1	2	1	1	1	1	1	1	1
S	3	3	2	3	3	2	2	3	3	3	3	3	3
T	2	1	2	1	2	2	1	2	2	2	2	2	2
Total	7	7	7	6	7	7	5	7	7	7	8	7	7
Marge de variat.	2	2	1	2	2	1	1	2	2	2	2	2	2
Total général	61 ²	58	60	59	63	65	59	62	64	63	66	64	63 ¹
Somme des marges	15 ⁴	19	15	17	14	13	14	16	14	14	14	13	14 ³

Différences entre les notes accordées et la moyenne des notes accordées à un même candidat

Nom du candidat	a	b	c	d	e
A	0	0	0	-1	0
B	0	0	0	0	1
C	1	0	0	0	0
D	0	-1	-1	0	0
Marge de variat.	1	1	1	1	1
E	0	0	0	1	0
F	0	0	0	0	0
G	-1	0	0	0	-1
H	-1	0	0	0	-1
Marge de variat.	1	0	0	1	1
I	-1	-1	-1	-1	1
J	0	-1	0	-1	1
K	0	0	0	1	0
L	-1	0	0	-1	0
Marge de variat.	1	1	1	2	1
M	0	-2	-1	0	-1
N	0	-1	0	0	0
O	1	0	0	-1	0
P	0	1	0	0	0
Marge de variat.	1	3	1	1	1
Q	0	1	0	0	0
R	0	0	1	0	0
S	0	0	-1	0	0
T	0	-1	0	-1	0
Marge de variat.	0	2	2	1	0
Somme des marges ⁵	4	7	5	6	4

Contrôles.

Les règles suivantes n'ont pas été appliquées, G. Peaker s'est inspiré de règles de qualité utilisées dans l'industrie. dans le domaine des examens

Le contrôle ainsi réalisé est insuffisant mais non très fin. Aussi, si malheureusement un aspect de la correction paraît être un problème !

1° Sévérité.

L'expérience révèle que le problème vient porter surtout sur ce point. Cependant, la solution est heureusement

est-il classé de la même façon
révélée par la corrélation entre

modérateurs, à ces trois points
abord un même paquet de 20

s, toujours au hasard, en cinq
groupe, les élèves sont classés

s opérations suivantes, dans
bleau :

ENTRE LES MODERATEURS

par 12 modérateurs

MODERATEURS								Moyenne
f	g	h	i	j	k	l		
4	3	1	4	4	4	4	4	4
2	1	1	2	2	2	2	2	2
5	5	5	5	5	5	5	5	5
3	3	2	2	3	2	3	2	2
14	12	9	13	14	13	14	13	13
3	4	4	3	3	3	3	3	3
5	3	3	3	3	3	3	3	3
1	1	1	1	1	1	1	1	1
3	6	6	6	6	6	6	6	6
4	3	4	4	2	4	4	4	4
15	13	14	14	12	14	14	14	14
4	5	5	5	5	5	5	5	5
3	2	3	3	3	3	3	3	3
2	2	2	2	2	2	2	2	2
5	5	6	5	5	5	4	5	5
2	2	3	2	2	3	2	2	2
12	11	14	12	12	13	11	12	12
3	3	4	3	3	3	2	3	3
4	5	5	5	4	5	4	4	4
5	4	4	4	4	4	4	4	4
3	4	4	4	5	4	5	4	4
5	5	5	5	5	5	5	5	5
17	18	18	18	18	18	18	17	17
2	1	1	1	1	1	1	1	1
1	1	1	1	1	2	1	1	1
2	1	1	1	1	1	1	1	1
2	2	3	3	3	3	3	3	3
2	1	2	2	2	2	2	2	2
7	5	7	7	7	8	7	7	7
1	1	2	2	2	2	2	2	2
65	59	62	64	63	66	64	63	63
13	14	16	14	14	14	13	14	14

Différences entre les notes accordées par chaque modérateur
et la moyenne des notes accordées par tous les modérateurs
à un même candidat.

Nom du candidat	MODERATEURS											
	a	b	c	d	e	f	g	h	i	j	k	l
A	0	0	0	-1	0	0	-1	-3	0	0	0	0
B	0	0	0	0	1	0	-1	-1	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0	0	0
D	0	-1	-1	0	0	1	1	0	0	1	0	1
Marge de variat.	1	1	1	1	1	1	2	3	0	1	0	1
E	0	0	0	1	0	2	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0
G	-1	0	0	0	-1	-1	0	0	0	0	0	0
H	-1	0	0	0	-1	0	-1	0	0	-2	0	0
Marge de variat.	1	0	0	1	1	3	1	0	0	2	0	0
I	-1	-1	-1	-1	1	0	-1	0	0	0	0	0
J	0	-1	0	-1	1	0	0	0	0	0	0	0
K	0	0	0	1	0	0	0	1	0	0	0	-1
L	-1	0	0	-1	0	0	0	1	0	0	1	0
Marge de variat.	1	1	1	2	1	0	1	1	0	0	1	1
M	0	-2	-1	0	-1	0	1	1	1	0	1	0
N	0	-1	0	0	0	1	0	0	0	0	0	0
O	1	0	0	-1	0	-1	0	0	0	1	0	1
P	0	1	0	0	0	0	0	0	0	0	0	0
Marge de variat.	1	3	1	1	1	2	1	1	1	1	1	1
Q	0	1	0	0	0	0	0	0	0	0	1	0
R	0	0	1	0	0	1	0	0	0	0	0	0
S	0	0	-1	0	0	-1	-1	0	0	0	0	0
T	0	-1	0	-1	0	0	-1	0	0	0	0	0
Marge de variat.	0	2	2	1	0	2	1	0	0	0	1	0
Somme des marges ⁵	4	7	5	6	4	8	6	5	1	4	3	3

Contrôles.

Les règles suivantes n'ont rien de magique. Pour les formuler, G. Peaker s'est inspiré des techniques de contrôle de qualité utilisées dans l'industrie. Des expériences très poussées dans le domaine des examens ont permis de les adapter.

Le contrôle ainsi réalisé est raisonnable, expéditif, efficace, mais non très fin. Aussi, si malgré la grossièreté des critères, un aspect de la correction paraît inacceptable, il y a certainement un problème !

1° Sévérité.

L'expérience révèle que les efforts d'harmonisation doivent porter surtout sur ce point. On va le voir, en cas de problème, la solution est heureusement facile.

Règle :

- a) Partir de la moyenne¹ des totaux généraux (ici 63) ;
- b) Ne pas tolérer des écarts au-delà d'une marge centrale de 10 points¹ par rapport à cette moyenne, soit 5 points en plus ou en moins (ici : 58-68).

Constatation :

Tous les totaux généraux sont compris dans cette marge.
Tous les correcteurs sont donc d'une sévérité acceptable.

Solution en cas d'excès de sévérité ou de générosité :
Supposons que l'on se trouve devant les totaux suivants :
75-72-68-66-65-64-60-49-46-44-41-40.

On constate que si on réunit le plus sévère et le plus généreux, puis le deuxième plus sévère et le deuxième plus généreux, etc., les moyennes des couples sont fort proches de la moyenne générale.

Soit :			Moyenne
(1)	40	75	57,5
(2)	41	72	56,5
(3)	44	68	56
(4)	46	66	56
(5)	49	65	57
(6)	60	64	62
Moyenne générale			57,5
Marge de 10 points			52,5-62,5

Seuls, les deux correcteurs du couple n° 6 se situent entre 52,5 et 62,5. Ces deux correcteurs seront autorisés à corriger seuls. Dans les autres cas, les copies devront être vues par les deux membres d'un des groupes constitués, et la moyenne des points sera faite.

2° Discrimination.

Règle :

Le total des marges moyennes (3) ne peut pas être supérieur au double du total des marges d'un correcteur (4) et réciproquement.

Constatation :

Aucun problème.

¹ Les correcteurs notent 20 compositions de 1 à 5. Dix points équivalent à une différence moyenne d'un demi-point par copie pour l'ensemble des 20 copies.

3° Conformité.

Règle :

Le total des marges des différen-
ciées par un modérateur et la
par l'ensemble ne doit pas être su-

Constatation :

Aucun problème.

Conclusion.

Dans le cas présent, tous
les trois épreuves de contrôle. Ils

Où en sommes-nous ? Des d
un échantillon est maintenant
notateurs l'ont vu !).

Comme chaque modérateur
des opérations ira vite.

D. Nouvelle correction des écha

Les opérations sont pratique
contrôle des modérateurs. Nous
détaillé parce que la présentatio
rateur et un seul professeur) d
semble.

¹ Nous montrons p. 167 que l'on obtie
estimation de la corrélation entre deux séries d

... totaux généraux (ici 63) ;
 ... au-delà d'une marge centrale de
 ... cette moyenne, soit 5 points en
 ... 8).

... compris dans cette marge.
 ... d'une sévérité acceptable.

... de sévérité ou de générosité :
 ... devant les totaux suivants :
 ... 41-40.

... le plus sévère et le plus généreux,
 ... et le deuxième plus généreux,
 ... sont fort proches de la moyenne

	Moyenne
75	57,5
72	56,5
68	56
66	56
65	57
64	62
.....	57,5
.....	52,5-62,5

... couple n° 6 se situent entre 52,5
 ... seront autorisés à corriger seuls.
 ... devront être vues par les deux
 ... constitués, et la moyenne des

(3) ne peut pas être supérieur
 d'un correcteur (4) et récipro-

... de 1 à 5. Dix points équivalent à une
 ... ie pour l'ensemble des 20 copies.

3° Conformité.

Règle :

Le total des marges des différences (5) entre les notes attri-
 buées par un modérateur et la moyenne des notes attribuées
 par l'ensemble ne doit pas être supérieur à 12¹.

Constatation :

Aucun problème.

Conclusion.

Dans le cas présent, tous les modérateurs ont surmonté
 les trois épreuves de contrôle. Ils pourront donc travailler seuls.

Où en sommes-nous ? Des douze échantillons de 20 copies,
 un échantillon est maintenant corrigé définitivement (douze
 notateurs l'ont vu !).

Comme chaque modérateur peut travailler seul, la suite
 des opérations ira vite.

D. Nouvelle correction des échantillons restants et contrôle.

Les opérations sont pratiquement les mêmes que pour le
 contrôle des modérateurs. Nous donnons, toutefois, un exemple
 détaillé parce que la présentation plus concise (un seul modé-
 rateur et un seul professeur) donne une meilleure vue d'en-
 semble.

¹ Nous montrons p. 167 que l'on obtient avec ce repère empirique une bonne estimation de la corrélation entre deux séries de notes.

**CALCULS POUR LA COMPARAISON
ENTRE UN PROFESSEUR ET UN MODERATEUR
(1 est la meilleure note ; 5 la moins bonne !)**

5 groupes de 4	Elèves choisis au hasard	Note attribuée par				Différence (Modérateur-Professeur)
		Modérateur		Professeur		
1 ^{er} groupe	Henri	Min.	5	5	Min.	0
	Jean		4	4		0 Min.
	Paul	Max.	2	1	Max.	1
	Pierre		3	2		1 Max.
	Marge : Min-Max	3	14	12	4	Marge des différences = 1
2 ^e groupe	André	Max.	2	1	Max.	1 Max.
	Edouard		1	1		0 Min.
	Jules	Min.	1	1	Min.	0
	René		4	4		0
	Marge : Min-Max	3	8	7	3	Marge des différences = 1
3 ^e groupe	Antoine		3	3	Max.	0 Max.
	Camille		5	5		0
	Eugène	Max.	2	3		-1 ¹ Min.
	Jérôme	Min.	6	6	Min.	0
	Marge : Min-Max	4	16	17	3	Marge des différences = 1
4 ^e groupe	Jacques	Min.	5	5	Min.	0
	Laurent	Max.	1	1	Max.	0 Min.
	Martin		4	3		1 Max.
	Victor		4	3		1
	Marge : Min-Max	4	14	12	4	Marge des différences = 1
5 ^e groupe	Bruno	Max.	3	2	Max.	1
	Hugues	Min.	4	2		2 Max.
	Léon		4	5	Min.	-1 Min.
	Simon		3	2		1
	Marge : Min-Max	1	14	11	3	Marge des différences = 3
Les cinq groupes réunis	Total des grades		66	59		Total des marges de différences = 7
	Marges réunies	15	(3)	(4)	17	(5)

* On considère que (0) est plus grand que (-1).

Contrôles.

1° Sévérité des correcteurs

La différence entre total (3) et total (4) est supérieure à 10.

Ici, 66 — 59 = 7 : Acceptable

2° Discrimination.

Le total (3) ne peut pas être inférieur à 10 (4) et réciproquement.

Ici, 15 et 17 : Acceptable.

3° Conformité.

Le total des marges des différences doit être supérieur à 12.

Ici, = 7 : acceptable¹.

4° Conclusion.

Dans ce cas, aucun problème de concordance des notes du professeur sont donc terminées.

Si les notes du professeur et du modérateur sont en accord, l'idéal est d'engager une discussion pour parvenir à un accord.

Ce n'est pas toujours possible en temps.

La solution suivante est proposée : si le professeur constate un problème, il invite le modérateur à juger indépendamment le même élève. Le modérateur calcule la moyenne entre les notes et le professeur est invité à ajuster toutes ses notes en fonction des notes du modérateur.

Selon G. Peaker, les corrections doivent devenir de plus en plus rapides et la modération fonctionne, en particulier dans les cas où les notes élevées qu'ils provoquent chaque fois.

¹ Cette façon simple de procéder évite les problèmes de deux séries de notes. On verra d'ailleurs que l'on obtient le même coefficient calculé par la méthode classique.

LA COMPARAISON
 R ET UN MODERATEUR
 e ; 5 la moins bonne !)

Note attribuée par		Professeur	Différence (Modérateur- Professeur)
Modérateur	Professeur		
5 4 2 3	5 4 1 2	Min. Max.	0 0 Min. 1 1 Max.
14	12	4	Marge des différences = 1
2 1 1 4	1 1 1 4	Max. Min.	1 Max. 0 Min. 0 0
8	7	3	Marge des différences = 1
3 5 2 6	3 5 3 6	Max. Min.	0 Max. 0 -1 ¹ Min. 0
16	17	3	Marge des différences = 1
5 1 4 4	5 1 3 3	Min. Max.	0 0 Min. 1 Max. 1
14	12	4	Marge des différences = 1
3 4 4 3	2 2 5 2	Max. Min.	1 2 Max. -1 Min. 1
14	11	3	Marge des différences = 3
66 (1) (3)	59 (2) (4)	17	Total des marges de différences = 7 (5)

1 (-1).

Contrôles.

1° Sévérité des correcteurs.

La différence entre total (1) et total (2) ne doit pas être supérieure à 10.

Ici, 66 — 59 = 7 : Acceptable.

2° Discrimination.

Le total (3) ne peut pas être supérieur au double du total (4) et réciproquement.

Ici, 15 et 17 : Acceptable.

3° Conformité.

Le total des marges des différences (5) ne doit pas être supérieur à 12.

Ici, = 7 : acceptable¹.

4° Conclusion.

Dans ce cas, aucun problème ne semble se poser. Les notes du professeur sont donc acceptées et les opérations sont terminées.

Si les notes du professeur ne semblent pas acceptables, l'idéal est d'engager une discussion avec lui afin de rechercher un accord.

Ce n'est pas toujours possible, en particulier faute de temps.

La solution suivante est plus expéditive. Si un modérateur constate un problème, il invite quatre de ses collègues à recorriger indépendamment le même échantillon de vingt copies. On calcule la moyenne entre les cinq modérateurs, et le professeur est invité à ajuster toutes ses notes en fonction de celles des modérateurs.

Selon G. Peaker, les cas nécessitant pareille correction deviennent de plus en plus rares, à mesure que le système de modération fonctionne, en raison des échanges de vues fructueux qu'ils provoquent chaque fois.

¹ Cette façon simple de procéder évite le long calcul de la corrélation entre les deux séries de notes.
 On verra d'ailleurs que l'on obtient un coefficient de corrélation très proche du coefficient calculé par la méthode classique en faisant l'opération :

$$\frac{152 + 172 - 72}{2 \times 15 \times 17} = .90$$

E. Comment ajuster des notes discordantes ?

Il s'agit, soit d'un simple ajustement de la moyenne (ou du médian) par relèvement ou abaissement des notes de tous les élèves (sévérité ou générosité excessives), soit aussi d'une modification de la distribution générale, pour améliorer la discrimination.

1° Ajustement du médian.

Pour un même examen, la situation est la suivante :

	Quartile supérieur	Médian	Quartile inférieur	Ecart type approximatif
Modérateur	16	13	11	$\sigma = 4$
Professeur	11	8	6	$\sigma = 4$

Le professeur est plus sévère que le modérateur, mais il discrimine aussi bien. Si le professeur ajoute 5 points à chaque élève, le parallélisme est rétabli.

2° Ajustement du médian et de l'écart type.

Situation de départ :

	Quartile supérieur	Médian	Quartile inférieur	Ecart type approximatif
Modérateur	16	13	11	$\sigma = 4$
Professeur	16	12	9	$\sigma = 5$

Pour aligner sur le modérateur, on corrige d'abord pour le médian ; un point est ajouté à tous les élèves.

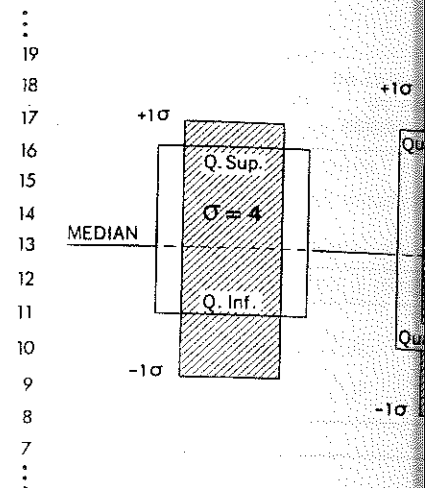
Supposons que les notes de départ du professeur sont :

... 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 ...

|
médian

¹ Adapté d'après D. MATHER et al., o.c., p. 183.

Nous allons d'abord montrer l'ajustement de l'écart type.



I. Notes du modérateur.

Pour l'opération III, l'ajustement de l'écart type est un peu plus complexe. La confection d'une table de conversion est nécessaire.

Pour trouver la nouvelle note, il suffit d'ajouter le nouveau sig. $13 + 4 = 17$. La nouvelle note est $17 - 4 = 9$, etc.

discordantes ?

justement de la moyenne (ou du
baissement des notes de tous les
à excessives), soit aussi d'une
générale, pour améliorer la discri-

situation est la suivante :

Médian	Quartile inférieure	Ecart type approximatif
13	11	$\sigma = 4$
8	6	$\sigma = 4$

ère que le modérateur, mais il
seigneur ajoute 5 points à chaque

de l'écart type.

Médian	Quartile inférieure	Ecart type approximatif
13	11	$\sigma = 4$
12	9	$\sigma = 5$

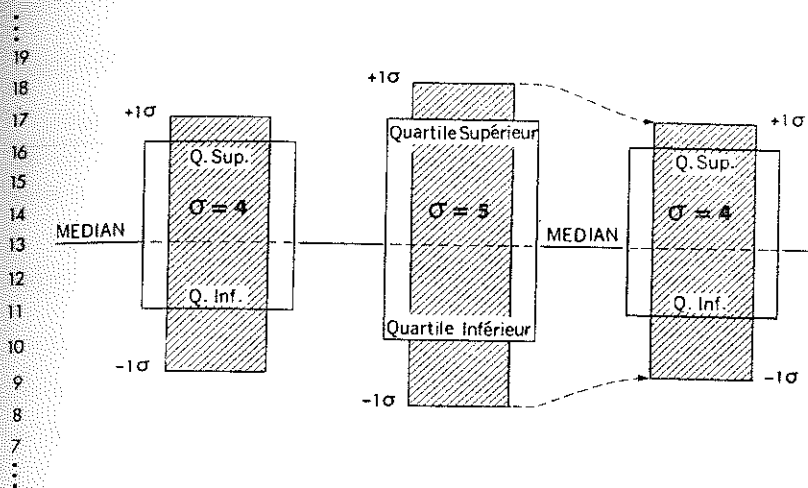
ur, on corrige d'abord pour le
us les élèves.

départ du professeur sont :

12 11 10 9 8 7 6 5 ...

dian

Nous allons d'abord montrer par le dessin en quoi consiste
l'ajustement de l'écart type.



I. Notes du modérateur.

II. Notes du professeur, augmentées d'un point. Le médian est devenu le même que celui du modérateur. Reste à ajuster σ .

III. Notes du professeur diminuées une seconde fois pour aligner l'écart type sur celui du modérateur.

Pour l'opération III, l'ajustement en fonction du nouvel écart type est un peu plus compliqué que pour le médian. La confection d'une table de conversion facilitera les opérations.

Pour trouver la nouvelle note correspondant à $+ 1 \sigma$, il suffit d'ajouter le nouveau sigma au nouveau médian, soit $13 + 4 = 17$. La nouvelle note correspondant à $- 1 \sigma = 13 - 4 = 9$, etc.

Pour les notes correspondant à des fractions de sigma, on calcule d'abord l'écart par rapport au médian, on multiplie par 4/5 et on ajoute le résultat au médian. L'exemple suivant va éclairer cette phase.

	Notes de départ du professeur	Notes ajustées en fonction du nouveau médian	Second ajustement en fonction du nouvel écart type (σ)
	.	.	.
	19	.	.
	18	.	.
+ 1 σ	17	18	17 *
	16	17	16 **
	15	16	15 ***
	14	15	15 ****
	13	14	14
Médian	12	13	13
	11	12	12
	10	11	11
	9	10	11
	8	9	10
- 1 σ	7	8	9
	6	.	.
	5	.	.
	.	.	.

Comment a-t-on trouvé ces chiffres ?

* Nouveau médian : 13
 $13 + 4 = 17$
 Nouvel écart type : 4

** Dans la 2^e colonne, 17 est à 4 points du médian.

Multiplier 4 par $\frac{4}{5} = 3,2$

Médian + 3,2 = 16,2 arrondi à 16

*** $16 - 13 = 3$; $3 \times \frac{4}{5} = 2,4$; $13 + 2,4 = 15,4$ arrondi à 15

**** $15 - 13 = 2$; $2 \times \frac{4}{5} = 1,6$; $13 + 1,6 = 14,6$ arrondi à 15

Problème :

On veut classer les élèves
 fin de leurs études. A sera
 sera le grade inférieur et vau

On veut tenir compte d
 selon les mêmes échelles à c

— travail de l'année = T.A.

— travaux pratiques = T.P.

— test régional = T.R.

Le test régional se voit
 les deux autres éléments. D'o

nt à des fractions de sigma,
pport au médian, on multiplie
au médian. L'exemple suivant

stées en n du médian	Second ajustement en fonction du nouvel écart type (σ)
.	.
.	.
.	.
.	.
.	17 *
.	16 **
.	15 ***
.	15 ****
.	14
.	13
.	12
.	11
.	11
.	10
.	9
.	.
.	.
.	.

chiffres ?

$$13 + 4 = 17$$

à 4 points du médian.

i à 16

$$; 13 + 2,4 = 15,4 \text{ arrondi à } 15$$

$$; 13 + 1,6 = 14,6 \text{ arrondi à } 15$$

VI. LA NOTE DE FIN D'ANNEE.

Travail de l'année + travaux pratiques + test.

Problème ¹

On veut classer les élèves en cinq groupes de mérite à la fin de leurs études. A sera le grade supérieur et vaudra 1, E sera le grade inférieur et vaudra 5.

On veut tenir compte de trois éléments, évalués chacun selon les mêmes échelles à cinq degrés :

- travail de l'année = T.A.
- travaux pratiques = T.P.
- test régional = T.R.

Le test régional se voit attribuer autant d'importance que les deux autres éléments. D'où la pondération :

$$T.A. = \frac{1}{4}$$

$$T.P. = \frac{1}{4}$$

$$T.R. = \frac{1}{2}$$

¹ D'après O. MATHER et al., o.c., pp. 149-154.

Voici le tableau général des grades et le tableau des résultats après pondération et ajustement final.

ATTRIBUTION DU GRADE FINAL

Elève N°	Test régional T.R		Evaluation par l'école		Pondération			Total	Grade final
	Score	Note	T.A.	T.P.	T.R $\times \frac{1}{2}$	T.A $\times \frac{1}{4}$	T.P $\times \frac{1}{4}$		
1	59	2	3	3	1	1	$\frac{3}{4}$	$2\frac{3}{4}$	3
2	77	1	1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1	1
3	51	3	4	3	$1\frac{1}{2}$	1	$\frac{3}{4}$	$3\frac{1}{4}$	3
4	12	6	5	5	3	$1\frac{1}{4}$	$1\frac{1}{4}$	$5\frac{1}{2}$	6
5	53	3	3	4	$1\frac{1}{2}$	1	1	$3\frac{1}{2}$	4
6	40	4	2	5	2	$\frac{3}{4}$	$1\frac{1}{4}$	4	4
7	66	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
8	60	2	3	4	1	$\frac{3}{4}$	1	$2\frac{3}{4}$	3
9	38	5	4	5	$2\frac{1}{2}$	$1\frac{1}{4}$	$1\frac{1}{4}$	5	5
10	70	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
11	56	2	2	3	1	$\frac{1}{2}$	$\frac{3}{4}$	$2\frac{1}{4}$	2
12	69	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
13	44	4	4	5	2	1	$1\frac{1}{4}$	$4\frac{1}{4}$	4
14	64	2	1	2	1	$\frac{1}{4}$	$\frac{1}{2}$	$1\frac{3}{4}$	2
15	19	6	5	5	3	$1\frac{1}{4}$	$1\frac{1}{4}$	$5\frac{1}{2}$	6
16	49	3	2	4	$1\frac{1}{2}$	$\frac{3}{4}$	1	$3\frac{1}{4}$	3
17	54	3	2	1	$1\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$2\frac{1}{4}$	2
18	47	4	1	2	2	$\frac{1}{4}$	$\frac{1}{2}$	$2\frac{3}{4}$	3
19	52	3	3	1	$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$	$2\frac{1}{2}$	2
20	48	3	3	3	$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$	3	3
21	50	3	1	2	$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$2\frac{1}{4}$	2
22	24	6	3	4	3	$\frac{3}{4}$	1	$4\frac{3}{4}$	5
23	61	2	4	2	1	1	$\frac{1}{2}$	$2\frac{1}{2}$	2
24	57	2	1	3	1	$\frac{1}{4}$	$\frac{3}{4}$	2	2
25	42	4	3	5	2	$\frac{3}{4}$	$1\frac{1}{4}$	4	4
26	35	5	5	2	$2\frac{1}{2}$	$1\frac{1}{4}$	$\frac{3}{4}$	$4\frac{1}{2}$	4
27	45	4	4	1	2	1	$\frac{1}{4}$	$3\frac{1}{4}$	3
28	41	4	2	4	2	$\frac{3}{4}$	1	$3\frac{3}{4}$	4
29	27	5	5	5	$2\frac{1}{2}$	$1\frac{1}{4}$	$1\frac{1}{4}$	5	5
30	43	4	3	2	2	1	$\frac{1}{2}$	$3\frac{1}{2}$	4
31	67	1	1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1	1
32	31	5	5	3	$2\frac{1}{2}$	$1\frac{1}{4}$	1	$4\frac{3}{4}$	5
33	72	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1

Les notes pondérées en italiques seront ajustées en fonction du test régional (T.R.).

La lecture du tableau nous donne trois notes pondérées : 1, 1, 1 final 3*. L'élève 3 : $1\frac{1}{2} + 1 + 1 = 3\frac{1}{2}$

Pourquoi un dernier ajustement du grade final ? Parce que la nouvelle concentration vers le haut tel que, dans une échelle finale, risque de se retrouver au milieu.

L'exemple fictif suivant :

ELEVES		Te
Pierre		
Paul		
Jean		
Roger		
Henri		
Dispersion		

Si l'on se reporte au tableau des grades pondérés, on rencontre (à part l'élève 3, qui est tout), alors qu'on n'en retrouve pas la *Total*. Les notes d'excellence sont donc ajustées vers le bas des grades.

Le procédé suivant assure l'équivalence entre les élèves d'une même région. On ajuste le grade final sur le niveau régional en tenant compte de l'état étant reconnu comme l'étalon.

* Les chiffres en italiques dans le tableau ci-dessus sont l'accord de l'école, ajusté la note du test régional dans la région.

grades et le tableau des résultats final.

GRADE FINAL

Pondération			Total	Grade final
T.R $\times \frac{1}{2}$	T.A $\times \frac{1}{4}$	T.P $\times \frac{1}{4}$		
1 $\frac{1}{2}$	<i>1</i> $\frac{1}{4}$	$\frac{3}{4}$ $\frac{1}{4}$	$2\frac{3}{4}$ 1	3 1
$1\frac{1}{2}$	1	$\frac{3}{4}$	$3\frac{1}{4}$	3
3 $1\frac{1}{2}$	$1\frac{1}{4}$ <i>1</i>	$1\frac{1}{4}$ 1	$5\frac{1}{2}$ $3\frac{1}{2}$	6 4
2 $\frac{1}{2}$	$\frac{3}{4}$ $\frac{1}{2}$	$1\frac{1}{4}$ $\frac{1}{2}$	4 $1\frac{1}{2}$	4 1
1 $2\frac{1}{2}$	$\frac{3}{4}$ $1\frac{1}{4}$	1 $1\frac{1}{4}$	$2\frac{3}{4}$ 5	3 5
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1
1 $\frac{1}{2}$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{3}{4}$ $\frac{1}{2}$	$2\frac{1}{4}$ $1\frac{1}{2}$	2 1
2 1	1 $\frac{1}{4}$	$1\frac{1}{4}$ $\frac{1}{2}$	$4\frac{1}{4}$ $1\frac{3}{4}$	4 2
3	$1\frac{1}{4}$	$1\frac{1}{4}$	$5\frac{1}{2}$	6
$1\frac{1}{2}$	$\frac{3}{4}$	1	$3\frac{1}{4}$	3
$1\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$2\frac{1}{4}$	2
2	$\frac{1}{4}$	$\frac{1}{2}$	$2\frac{3}{4}$	3
$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$	$2\frac{1}{2}$	2
$1\frac{1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$	3	3
$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	$2\frac{1}{4}$	2
3	$\frac{3}{4}$	1	$4\frac{3}{4}$	5
1	1	$\frac{1}{2}$	$2\frac{1}{2}$	2
1	$\frac{1}{4}$	$\frac{3}{4}$	2	2
2	$\frac{3}{4}$	$1\frac{1}{4}$	4	4
$2\frac{1}{2}$	$1\frac{1}{4}$	$\frac{3}{4}$	$4\frac{1}{2}$	4
2	1	$\frac{1}{4}$	$3\frac{1}{4}$	3
2	$\frac{3}{4}$	1	$3\frac{3}{4}$	4
$2\frac{1}{2}$	$1\frac{1}{4}$	$1\frac{1}{4}$	5	5
2	1	$\frac{1}{2}$	$3\frac{1}{2}$	4
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1	1
$2\frac{1}{2}$	$1\frac{1}{4}$	1	$4\frac{3}{4}$	5
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	1

ques seront ajustées en fonc-

La lecture du tableau nous montre que l'élève 1 obtient les trois notes pondérées : 1, 1 et $\frac{3}{4}$. Soit au total $2\frac{3}{4}$; grade final 3^* . L'élève 3 : $1\frac{1}{2} + 1 + \frac{3}{4} = 3\frac{1}{4}$; grade final 3.

Pourquoi un dernier ajustement au moment d'attribuer le grade final ? Parce que l'addition des grades provoque une nouvelle concentration vers la moyenne ; le rétrécissement est tel que, dans une échelle finale à 5 degrés, toute la population risque de se retrouver au milieu.

L'exemple fictif suivant illustre ce phénomène.

ELEVES	GRADE		
	Test R.	Trav. année	Moyenne
Pierre	1	5	3
Paul	2	4	3
Jean	3	3	3
Roger	4	2	3
Henri	5	1	3
Dispersion	4	4	0

Si l'on se reporte au tableau p. 172, on observe que, parmi les grades pondérés, on rencontre un grand nombre de 1 (17 en tout), alors qu'on n'en retrouve plus que deux dans la colonne *Total*. Les notes d'excellence ont été noyées par l'addition des grades.

Le procédé suivant assure une meilleure justice distributive entre les élèves d'une même classe et un meilleur alignement sur le niveau régional révélé par le test régional, celui-ci étant reconnu comme l'étalon le plus sûr.

* Les chiffres en italiques dans le tableau indiquent que le modérateur a, avec l'accord de l'école, ajusté la note du T.A. afin de l'harmoniser avec le niveau moyen dans la région.

On part du tableau suivant.

TEST REGIONAL		TOTAL DES GRADES PONDERES		GRADE FINAL	
Grade (1)	N. d'élèves ayant obtenu ce grade (2)	Grade (3)	N. d'élèves ayant obtenu ce grade (4)	N. d'élèves auxquels il est attribué (5)	Grade (9)
1		1			1
		1 1/4			
		1 1/2			
		1 3/4			
2		2			2
		2 1/4			
		2 1/2			
		2 3/4			
3		3			3
		3 1/4			
		3 1/2			
		3 3/4			
4		4			4
		4 1/4			
		4 1/2			
		4 3/4			
5		5			5
Au-delà		Au-delà			Non classé

Opérations.

1. Pointer dans la colonne (2) le nombre d'élèves et faire le total.
2. Commencer à pointer dans la colonne (4) et s'arrêter quand le total égale celui de la colonne (2). On pratique la première coupure à cet endroit. Afin de ne pas désavantager certains élèves, on dépasse, au besoin, le nombre de la colonne (3), de façon à épuiser le niveau où l'on s'est arrêté. On verra, par exemple, dans le tableau ci-dessous, qu'on accorde le grade 2 à 7 élèves et non à 6, parce que 2 d'entre eux ont obtenu 2 1/2. On ne pouvait évidemment attribuer un grade 2 à l'un et un grade 3 à l'autre.

TEST REGIONAL			DES GRADES
GRADE (1)	N. D'ÉLÈVES AYANT OBTENU CE GRADE (2)	GRADE (3)	N. D'ÉLÈVES AYANT OBTENU CE GRADE (4)
1	JHT I 6	1	
		1 1/4	
		1 1/2	
		1 3/4	
2	JHT I 6	2	
		2 1/4	
		2 1/2	
		2 3/4	
3	JHT II 7	3	
		3 1/4	
		3 1/2	
		3 3/4	
4	JHT II 7	4	
		4 1/4	
		4 1/2	
		4 3/4	
5	IIII 4	5	
AU-DELÀ	III 3	AU-DELÀ	

Conclusion.

Aucun des systèmes décrits ne mettraient d'améliorer notre situation.

Un premier choix s'opère entre la sélection ou comparaison. Ce choix attire notre attention.

Nous ne nous prononçons pas sur le choix. La décision appartient aux décideurs pédagogiques et aux enseignants.

Un effort intense d'expérience est nécessaire : petits groupes de travail, études nationales en collaboration avec les laboratoires sociaux et les laboratoires des universités.

A mesure que ces travaux se poursuivent, il faut faire émerger un nouveau système de notation et aussi des objectifs pédagogiques immédiatement, toutefois, la recherche pédagogique régionale nous y reviendra encore !

GRADES RES	GRADE FINAL	
	N. d'élèves auxquels il est attribué (5)	Grade (9)
d'élèves nt obtenu grade (4)		
		1
		2
		3
		4
		5
		Non classé

le nombre d'élèves et faire le
colonne (4) et s'arrêter quand
e (2). On pratique la première
ne pas désavantager certains
le nombre de la colonne (3),
l'on s'est arrêté.
le tableau ci-dessous, qu'on
es et non à 6, parce que 2
ne pouvait évidemment attri-
ade 3 à l'autre.

TEST REGIONAL		TOTAL DES GRADES PONDÉRÉS		GRADE FINAL	
GRADE (1)	N. D'ÉLÈVES AYANT OBTENU CE GRADE (2)	GRADE (3)	N. D'ÉLÈVES AYANT OBTENU CE GRADE (4)	N. D'ÉLÈVES AUXQUELS IL EST ATTRIBUÉ (5)	GRADE (6)
1	III I 6	1	// 2	6	1
		1 1/4	/// 4		
		1 1/2			
		1 3/4			
2	III I 6	2	/ 1	7	2
		2 1/4	/ 1		
		2 1/2	/// 3		
		2 3/4	// 2		
3	III II 7	3	/// 3	7	3
		3 1/4	/ 1		
		3 1/2	/// 3		
		3 3/4	// 2		
4	III II 7	4	/ 1	7	4
		4 1/4	/ 1		
		4 1/2	// 2		
		4 3/4	/ 1		
5	IIII 4	5	// 2	4	5
AU-DELÀ	III 3	AU-DELÀ		2	NON CLASSÉS

Conclusion.

Aucun des systèmes décrits n'est parfait, mais tous permettraient d'améliorer notre système traditionnel de notation.

Un premier choix s'opérera en fonction du but poursuivi : sélection ou comparaison. Ce second aspect a surtout retenu notre attention.

Nous ne nous prononçons pas en faveur d'un système particulier. La décision appartient aux autorités politiques et pédagogiques et aux enseignants.

Un effort intense d'expérimentation doit être fait à tous les niveaux : petits groupes de professeurs, recherches régionales et nationales en collaboration avec les centres psycho-médico-sociaux et les laboratoires de pédagogie expérimentale des universités.

A mesure que ces travaux avanceront, on verra probablement émerger un nouveau système tenant compte de nos traditions et aussi des objectifs particuliers à notre pays. Presque immédiatement, toutefois, la nécessité de puissants centres de recherche pédagogique régionaux se fera de nouveau sentir. Nous y reviendrons encore !

CINQUIE

LE MYTHE D
DE

CINQUIEME PARTIE

LE MYTHE DE LA COURBE
DE GAUSS

LE DANGEREUX MYTHE DE L

Dans les sciences humaines, la joue un rôle considérable, parce qu répartition de bien des aptitudes e moyens abondent, mais les génies les nains sont rares.

La courbe de Gauss est soit qui préside à notre naissance, soit d'un grand nombre de facteurs agis indépendante sur un individu ou u

Comme les tests mesurent sou de personnalité ou des performan est naturel qu'ils soient étalonnés ne : en gros, 70 % de moyens, 13 cres, 2 % d'excellents, 2 % de très

Au cours de la construction de ment les questions qui seraient ré de sujets. Le but poursuivi est attribuer la place qui lui revient. Bref, il s'agit d'organiser une sorte occupera nécessairement la premi

C'est pourquoi beaucoup de t res de connaissances sont d'exco tion.

Dans sa classe, l'enseignant ment différent. Son idéal n'est-il pa nent à lire, à calculer et, de façon tement toutes les connaissances par la société. *Instruire n'est pa* C'est s'efforcer que tous réussiss *la courbe de Gauss* prise comme

Les conséquences pédagogic particulièrement importantes.

LE DANGEREUX MYTHE DE LA COURBE DE GAUSS.

Dans les sciences humaines, la courbe en cloche de Gauss joue un rôle considérable, parce qu'elle est l'image même de la répartition de bien des aptitudes et des qualités : les individus moyens abondent, mais les génies et les idiots, les géants et les nains sont rares.

La courbe de Gauss est soit le reflet de la loi du hasard qui préside à notre naissance, soit la résultante de l'influence d'un grand nombre de facteurs agissant de façon plus ou moins indépendante sur un individu ou un objet.

Comme les tests mesurent souvent des aptitudes, des traits de personnalité ou des performances de vastes populations, il est naturel qu'ils soient étalonnés selon la répartition gaussienne : en gros, 70 % de moyens, 13 % de bons, 13 % de médiocres, 2 % d'excellents, 2 % de très mauvais.

Au cours de la construction de tels tests, on élimine notamment les questions qui seraient réussies par trop ou trop peu de sujets. Le but poursuivi est de classer chacun, de lui attribuer la place qui lui revient dans un groupe nombreux. Bref, il s'agit d'organiser une sorte de concours, où le plus fort occupera nécessairement la première place.

C'est pourquoi beaucoup de tests d'aptitudes ou d'inventaires de connaissances sont d'excellents *instruments de sélection*.

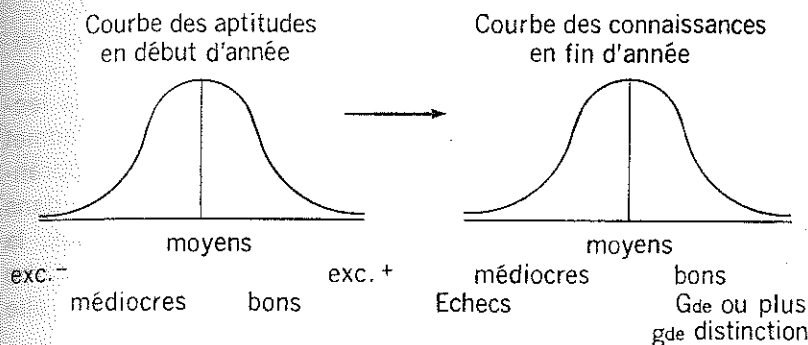
Dans sa classe, l'enseignant poursuit un objectif totalement différent. Son idéal n'est-il pas que *tous* les élèves apprennent à lire, à calculer et, de façon générale, à maîtriser parfaitement toutes les connaissances jugées nécessaires ou utiles par la société. *Instruire n'est pas sélectionner*. Au contraire ! C'est s'efforcer que *tous* réussissent. *C'est donc lutter contre la courbe de Gauss prise comme modèle de sélection*.

Les conséquences pédagogiques de ces observations sont particulièrement importantes.

Toutefois, une singulière distorsion se produit généralement dans les esprits. On considère cette répartition des aptitudes comme pronostique des résultats scolaires en fin d'année et on fixe le niveau de l'enseignement de telle façon que ce pronostic se vérifie : il sera « moyennement » difficile tout en permettant aux meilleurs de s'épanouir et en laissant une mince chance aux « médiocres ».

Insistons-y, la difficulté « moyenne » est déterminée par la moyenne des aptitudes du groupe considéré et non par une moyenne de difficulté *objective* des notions à enseigner¹.

Dans ces conditions, si le professeur fait *le même cours* à toute la classe, il est normal que la *courbe des connaissances* acquises en fin d'année respecte, à son tour, la distribution gaussienne.



La vocation de l'enseignement est-elle ainsi respectée ?

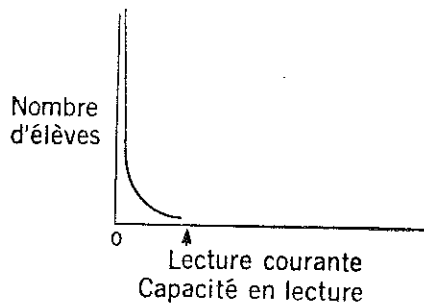
II. La courbe des connaissances.

Revenons au premier jour de l'année scolaire et, au lieu de considérer les aptitudes spécifiques des élèves, examinons leurs connaissances.

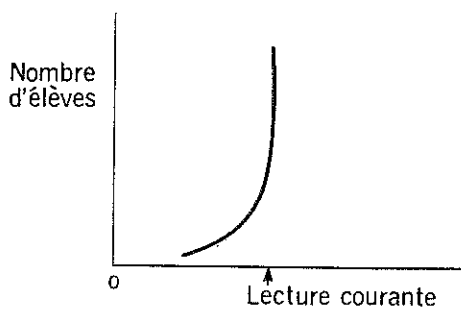
En toute logique, le rôle du professeur est de susciter l'apprentissage de connaissances *nouvelles*. Certes, imaginer que tous les individus formant une classe possèdent exactement la même quantité et qualité de connaissances est utopique. Néanmoins, le système de classes fixes que nous pratiquons repose sur l'hypothèse que *tous* se trouvent approximativement au même niveau. Sinon, comment oserions-nous encore dispenser le même enseignement à chacun ?

¹ Ainsi s'expliquent les différences considérables du niveau moyen selon les classes et les régions, dont nous avons parlé dans la première partie de ce livre.

Et en réalité ? Prenons le cas de l'entrée en première primaire. La majorité des enfants ne savent pas lire ; quelques-uns sont en bonne voie ; deux ou trois lisent déjà couramment. A ce moment, la courbe de la capacité en lecture épouse, en gros, la forme de la lettre *i*.

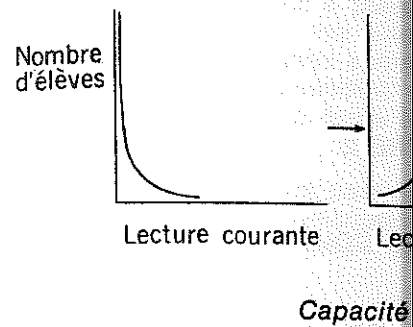


Or, bien que la courbe des aptitudes spécifiques à la lecture soit fort probablement gaussienne, l'instituteur n'admet pas d'emblée que, seule, une partie de la classe pourra apprendre à lire, au contraire. Pour autant que l'on ne verse pas dans le perfectionnisme, on peut dire qu'après un an ou deux, la grande majorité des élèves sauront lire couramment. La courbe des connaissances aura complètement changé de forme : elle ressemblera à un *j*.



Entre la courbe en *i* et la courbe en *j*, il est probable qu'un moment a existé où les mieux doués ont avancé le plus vite, où les moins doués ont traîné et où les moyens se sont situés entre les deux.

L'évolution est donc schém



Dans le cas de la lecture, c ou les responsables de l'ense d'une évolution différente.

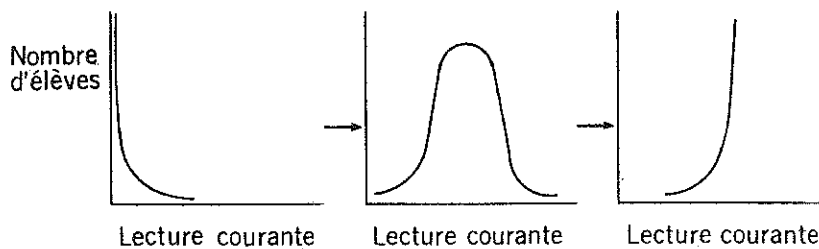
Mais, à partir de quel niv gence n'est-elle donc plus de m

cas de l'entrée en première pri-
me savent pas lire ; quelques-uns
trois lisent déjà couramment.
capacité en lecture épouse, en

aptitudes spécifiques à la lec-
tessienne, l'instituteur n'admet
rtie de la classe pourra appren-
ant que l'on ne verse pas dans
e qu'après un an ou deux, la
ont lire couramment. La courbe
tement changé de forme : elle

urbe en j , il est probable qu'un
oués ont avancé le plus vite,
où les moyens se sont situés

L'évolution est donc schématiquement la suivante :



Capacité en lecture.

Dans le cas de la lecture, on ne conçoit pas que les parents
ou les responsables de l'enseignement puissent se contenter
d'une évolution différente.

Mais, à partir de quel niveau de la scolarité, pareille exi-
gence n'est-elle donc plus de mise ?

CHAPITRE 2.

UNE PEDAGOGIE DE COURBE EN J.

Un examen où les résultats se distribuent selon la courbe en cloche de Gauss se prête bien à la sélection, au concours.

Or, pour trois raisons historiques principales, l'esprit de concours a imprégné notre enseignement pendant des siècles :

- 1° Pour des raisons socio-économiques, une partie seulement de la population scolarisable avait accès à l'école. Il y a cent ans, les familles modestes faisaient la première sélection en ne laissant finir l'école primaire qu'aux plus doués de leurs enfants, c'est-à-dire à ceux qui réussissaient le mieux leurs examens. Les bourses d'études, parcimonieusement distribuées, se gagnaient en concours.
 - 2° Jusqu'à ces derniers temps, — disons avant l'ère de l'ordinateur, — tous les pays industrialisés disposaient d'un énorme surplus de matière grise. On exploitait donc celle qui s'offrait à meilleur marché et avec le plus de facilité, c'est-à-dire que l'on se souciait surtout d'identifier les mieux doués. Même pour les enfants fortunés, l'enseignement secondaire jouait un rôle sélectif.
 - 3° Les connaissances psychologiques et pédagogiques étaient rudimentaires. Les maîtres n'étaient donc pas en mesure d'appliquer des traitements fins aux élèves éprouvant des difficultés d'apprentissage. D'ailleurs, aujourd'hui encore, on fait souvent répéter une année à l'élève qui n'a pas été capable des performances minima exigées pour le passage. Autrement dit, au lieu d'appliquer des remèdes particuliers, on se contente de placer de nouveau l'élève dans les conditions (même professeur, même méthode) dans lesquelles l'échec s'est produit.
- Chaque année scolaire étant considérée comme un filtre pour la suivante, une certaine quantité d'échecs paraissait donc normale. Bref, la répartition gaussienne semblait satisfaisante.

Mais le souci constant d'être le meilleur, causé, progressivement, une dévotion « latine ». On en est arrivé à penser que « comptent vraiment », les concours sont un « véritable jeu des idées abstraites » par une minorité possédant un degré élevé. Et l'on a forgé une conséquence.

Les élèves et leurs parents apprennent des choses sans grande discussion. L'enseignement « latin-mathématiques » se poursuit dans cette direction.

Assurément, on ne peut pas attendre de chacun un virtuose en matière de la littérature. Mais où se situe le degré l'accès à un savoir est déterminé par des aptitudes moyennes.

Par nature, chacun de nous a des domaines. Il ne manque pas de plus ou moins réfractaires à l'enseignement. On observe fréquemment qu'à force de faire redire les explications, de vouloir pour trouver une forme d'explication, les notions de mathématiques d'ailleurs accessibles sont bel et bien maîtrisées.

Pour autant qu'ils y consentent, les élèves moyens (c'est-à-dire, vu les études antérieures, probablement au niveau scolaire d'un niveau donné) peuvent qu'on ne l'imagine.

Dans cette perspective, la proposition de J. Carroll prend tout son sens : la quantité de temps demandée pour maîtriser une matière »¹.

L'implication de cette corrélation leur alloue le temps nécessaire normalement dans une classe pour obtenir un très bon résultat.

¹ J. CARROLL, A Model of School, 723-733.

COURBE EN J.

se distribuent selon la courbe
à la sélection, au concours.

ques principales, l'esprit de
nement pendant des siècles :

iques, une partie seulement
ait accès à l'école. Il y a cent
ent la première sélection en
qu'aux plus doués de leurs
réussissaient le mieux leurs
e, parcimonieusement distri-

disons avant l'ère de l'ordi-
nalisés disposaient d'un énor-
On exploitait donc celle qui
ec le plus de facilité, c'est-à-
d'identifier les mieux doués.
l'enseignement secondaire

es et pédagogiques étaient
aient donc pas en mesure
aux élèves éprouvant des

on fait souvent répéter une
capable des performances
ge. Autrement dit, au lieu
uliers, on se contente de
es conditions (même profes-
uelles l'échec s'est produit.

considérée comme un filtre
té d'échecs paraissait donc
enne semblait satisfaisante.

Mais le souci constant de la promotion des plus aptes a causé, progressivement, une déformation pédagogique plus subtile. On en est arrivé à penser que les connaissances « qui comptent vraiment », les connaissances « approfondies », le « véritable jeu des idées abstraites », ne sont assimilables que par une minorité possédant des aptitudes spécifiques à un degré élevé. Et l'on a forgé les méthodes d'enseignement en conséquence.

Les élèves et leurs parents acceptent d'ailleurs cet état de choses sans grande discussion. On s'inscrit rarement en section « latin-mathématiques » sans se sentir spécialement doué dans cette direction.

Assurément, on ne peut ambitionner de faire indifféremment de chacun un virtuose de la mathématique, du piano ou de la littérature. Mais où se situe la limite ? A partir de quel degré l'accès à un savoir est-il impossible à ceux qui possèdent des aptitudes moyennes, voire médiocres ?

Par nature, chacun de nous est médiocrement doué en bien des domaines. Il ne manque, par exemple, pas d'intellectuels plus ou moins réfractaires à la mathématique. Pourtant, on observe fréquemment qu'à force de vouloir, de persévérer, de faire redire les explications, de changer de manuel ou de maître pour trouver une forme d'enseignement qui convienne, des notions de mathématiques d'abord considérées comme inaccessibles sont bel et bien maîtrisées par certains.

Pour autant qu'ils y consacrent le temps nécessaire, les élèves moyens (c'est-à-dire, vu la sélection déjà opérée par les études antérieures, probablement plus de 80 % de la population scolaire d'un niveau donné) peuvent aller beaucoup plus loin qu'on ne l'imagine.

Dans cette perspective, la nouvelle définition de l'aptitude, proposée par J. Carroll prend toute sa valeur : « L'aptitude est la quantité de temps demandée par le *learner* pour dominer une matière »¹.

L'implication de cette conception est considérable : si on leur alloue le temps nécessaire, tous les élèves se trouvant normalement dans une classe pourraient arriver à un bon, voire un très bon résultat.

¹ J. CARROLL, A Model of School Learning, *Teachers College Record*, 1963, 64 : 723-733.

CHAPITRE 3.

LA THEORIE DE L'EVALUATION FORMATIVE.

L'expression *évaluation formative* — il l'oppose à *évaluation sommative* — a été forgée par Michael Scriven¹.

L'*évaluation sommative* (S. Roller parle dans ce cas de « norme psychotechnique ») nous est familière. Pour interpréter le score obtenu à un test classique d'inventaire de connaissances ou d'intelligence, on le situe dans une distribution statistique : la performance d'un individu est jugée par référence à celles d'autrui. De même, on classe souvent encore les élèves entre eux selon l'ordre croissant ou décroissant de leurs résultats scolaires, et c'est d'après la place ainsi occupée que bien des parents apprécient le travail de leurs enfants.

Or, dans les deux cas, le résultat est essentiellement relatif. Que le groupe de référence varie de composition et le résultat apparaît sous un autre jour.

Une simple différence d'âge peut aussi changer considérablement la face des choses. Dans bien des normes de tests de connaissances d'usage courant, un an d'âge ou une année scolaire en plus ou en moins suffisent pour qu'une même performance soit considérée comme médiocre ou bonne.

Que la connaissance soit acquise ou non n'a donc pas été la préoccupation première des constructeurs de ces tests, mais bien à quelle vitesse cette acquisition s'est réalisée. En nous reportant à la définition de J. Carroll, on évalue donc l'aptitude au lieu d'évaluer le contenu de l'apprentissage.

Les partisans de l'*évaluation formative* prennent le contre-pied de cette conception.

¹ M. SCRIVEN, *The Methodology of Evaluation*, in R. STAKER, (Ed.), *Perspectives of Curriculum Evaluation*, Chicago, Rand McNally, 1967.
Voir aussi : B.S. BLOOM, *Learning for Mastery*, in *Evaluation Comment*, 1968, 2. Notre exposé de la méthode de l'évaluation formative s'appuie directement sur cette publication. Plusieurs résultats d'expériences où la théorie du *mastery learning* a été appliquée sont publiés par J. BLOCK, B. BLOOM, HASTINGS (1970).

Puisque l'élève vient à l'école, n'est-il pas de le situer dans une norme ?

Imaginons qu'une analyse révèle qu'avec toutes ses nuances, l'élève participe passé, employé au cent cas différents. Pour chaque cas, il faut une maîtrise.

Selon le niveau scolaire, le résultat doit alors être fixé. Dans ce cas, le résultat est de nature absolue.

L'élève est noté en fonction du chemin parcouru dans l'acquisition.

En outre, si l'on ambitionne d'atteindre jusqu'à un niveau de connaissance totale, de la notion de maîtrise, le professeur enseignant est elle-même remis en question : d'appliquer indifféremment à tous les élèves pendant une même durée.

Que deviendrait l'enseignant ?

Trois problèmes cruciaux se posent :

1. Comment jalonner l'ascension ?
2. Comment conduire l'élève ?
3. Où se situe la limite de la progression ?

A ma connaissance, il n'y a pas de réponses complètes à ces questions. Pour l'instant, la seule réponse normale en sciences humaines est que nous devons apprendre à l'accepter. Notre discussion aborde maintenant des recherches à entreprendre ou à mener.

i. Jalonner l'ascension du savoir

En voyage, pour déterminer si l'on a atteint le but, deux conditions doivent être réunies : savoir où l'on est et où l'on veut aller. Une carte indiquant clairement le chemin est indispensable.

De même, en éducation, il faut définir les objectifs à atteindre et déterminer les conditions particulières qui y conduiront. Il faut savoir si c'est une affaire ou non à des apprécier.

Puisque l'élève vient à l'école pour apprendre, l'important n'est-il pas de le situer dans l'ascension du savoir ?

Imaginons qu'une analyse scientifique rigoureuse nous révèle qu'avec toutes ses nuances et ses complications, l'accord du participe passé, employé avec l'auxiliaire avoir, présente cent cas différents. Pour chacun, on peut définir des critères de maîtrise.

Selon le niveau scolaire, le nombre de cas à dominer peut alors être fixé. Dans ce contexte, l'évaluation scolaire change de nature.

L'élève est noté en fonction d'un critère objectif : le chemin parcouru dans l'acquisition.

En outre, si l'on ambitionne de conduire tous les élèves jusqu'à un niveau de connaissances minimum, sinon jusqu'à la maîtrise totale, de la notion ou de la technique, la façon d'enseigner est elle-même remise en cause : il n'est plus possible d'appliquer indifféremment à tous une seule et même méthode pendant une même durée.

Que deviendrait l'enseignement dans cette perspective ?

Trois problèmes cruciaux se posent :

1. Comment jalonner l'ascension du savoir ?
2. Comment conduire l'élève ?
3. Où se situe la limite pratique de cette pédagogie de la courbe en *j* ?

A ma connaissance, il n'existe pas encore de réponses complètes à ces questions. Pareille imprécision du savoir semble normale en sciences naturelles ou en médecine. Nous devons apprendre à l'accepter aussi dans les sciences de l'éducation. Notre discussion aboutit donc maintes fois sur des recherches à entreprendre ou à continuer.

1. Jalonner l'ascension du savoir.

En voyage, pour déterminer à quelle distance on se trouve du but, deux conditions doivent être remplies : d'une part, savoir où l'on est et où l'on va, et, d'autre part, disposer d'une carte indiquant clairement le chemin.

De même, en éducation, nous devons définir les objectifs à atteindre et déterminer avec précision les apprentissages particuliers qui y conduiront. Le problème varie selon que l'on a affaire ou non à des apprentissages de base. Pour ces der-

RE 3.

EVALUATION FORMATIVE.

ative — il l'oppose à évaluation — par Michael Scriven¹.

Roller parle dans ce cas de test familier. Pour interpréter le résultat d'un inventaire de connaissances dans une distribution statistique, on est jugé par référence à une courbe qui se souvient encore les élèves et qui décroît de leurs résultats. On se place ainsi occupée que bien de leurs enfants.

Le résultat est essentiellement relatif. On parle de composition et le résultat

peut aussi changer considérablement des normes de tests de l'un an d'âge ou une année passent pour qu'une même performance médiocre ou bonne.

La mesure ou non n'a donc pas été effectuée par les constructeurs de ces tests, mais la mesure s'est réalisée. En nous basant sur ce résultat, on évalue donc l'aptitude à l'apprentissage.

Les évaluations formative prennent le contre-

¹ Scriven, in R. STAKER, (Ed.), *Perspectives on Evaluation*, 1967.
² Scriven, in *Evaluation Comment*, 1968, 2. Notre s'appuie directement sur cette publication de *mastery learning* a été appliquée en France (1970).

niers, il faut découvrir l'enchaînement « critique » des matières, c'est-à-dire celui où l'une n'est accessible que si la précédente est assimilée. Pour cette raison, l'apprentissage des connaissances et des techniques de base doit, en dernière analyse, être linéaire, tandis que les acquisitions et les applications qui vont au-delà voient s'ouvrir devant elles des voies de plus en plus nombreuses.

Par exemple, quelle que soit la méthode d'enseignement, il n'est pas possible d'appliquer complètement une règle de trois, sans avoir — notamment — la notion de la multiplication et de la division. A un niveau plus élevé, comment faire du calcul intégral sans savoir ce qu'est une fonction ? Mais pareilles propositions sont encore trop vagues. Quels sont exactement les apprentissages nécessaires et suffisants pour pouvoir assimiler la règle de trois ? Et, parmi eux, lesquels sont critiques par rapport aux autres ?

La question est redoutable. Et il serait naïf de croire que la réflexion et les tâtonnements qui ont présidé à l'élaboration de nos meilleurs manuels scolaires l'ont résolue. Certes, le bon enchaînement des chapitres constitue une amorce de solution, mais il suffit d'essayer de programmer la moindre notion, ne fût-ce qu'en système skinnérien, pour découvrir toutes les ignorances, toute l'imprécision des méthodes d'enseignement actuelles.

Cependant, c'est précisément dans les efforts de programmation entrepris par des chercheurs et des enseignants de plus en plus nombreux que gît une des grandes sources de progrès.

Assurément, les sciences déductives sont le terrain d'élection de la méthode qui s'esquisse. Dussions-nous nous y borner, nous n'aurions pas accompli œuvre négligeable. Mais la limite n'apparaît pas si vite. Là où l'induction domine dans l'élaboration du savoir, la déduction réapparaît pour son application. D'ailleurs, l'apprentissage inductif aussi peut être systématiquement guidé.

II. Guider l'élève.

Une fois clairement défini le chemin qui conduit à un apprentissage, il faut y engager l'élève. Pour les acquisitions fondamentales, nous avons, en outre, décidé que *tous* devraient, en principe, arriver au but.

Dans ce cas, l'évaluation fréquente des progrès est essentielle. D'où la nécessité d'*exercices de maîtrise* et de *tests diagnostics* portant sur des matières très limitées et utilisés par les maîtres eux-mêmes.

Ces instruments — que gnants et des chercheurs pe suffisant — indiquent où l'élève et où il éprouve des difficultés

Les remèdes sont multiples pratique pédagogique quotidienne

- a) Indiquer de façon précise
- b) Le travail par sous-groupe élèves rencontre la ou les avantage à consacrer une groupe, les autres travail
- c) Si le groupe compte max gère que chacun d'eux so gêne de deux ou trois. C concurrence, mais luttent l'entraide joue.
- d) Dans les écoles comptant branche, un élève devrait explication, voire une cou professeur que le sien.
- e) Varier la présentation : m des manuels expliquant permettre le recours à moyens audio-visuels ; var
- f) En général, ne pas fixer mêmes choses.

Pareille pratique paie. B conduite dans cette ligne où p la meilleure note en fin d'ann gie de la courbe en j.

Peut-il toujours en être ai

III. Limite pratique du systèm

Même s'il est vrai que to capables d'atteindre un hau qu'on leur laisse un temps s d'une certaine limite, la duré

Comment permettre de pa l'enseignement secondaire ! C fois, pas remettre le systèm élèves en fonction de leurs ar vation et d'orientation, introd gnement secondaire de la plu bilité.

ement « critique » des matières,
accessible que si la précédente
n, l'apprentissage des connais-
se doit, en dernière analyse, être
ons et les applications qui vont
lles des voies de plus en plus

t la méthode d'enseignement, il
omplètement une règle de trois,
notion de la multiplication et
élève, comment faire du calcul
une fonction ? Mais pareilles
ues. Quels sont exactement les
ffisants pour pouvoir assimiler
x, lesquels sont critiques par

Et il serait naïf de croire que la
ont présidé à l'élaboration de
l'ont résolue. Certes, le bon
stitue une amorce de solution,
ammer la moindre notion, ne
pour découvrir toutes les igno-
s méthodes d'enseignement

dans les efforts de program-
rs et des enseignants de plus
grandes sources de progrès.

uctives sont le terrain d'élec-
Dussions-nous nous y borner,
re négligeable. Mais la limite
ction domine dans l'élabora-
paraît pour son application.
aussi peut être systématique-

chemin qui conduit à un
élève. Pour les acquisitions
e, décidé que *tous* devraient,

ente des progrès est essen-
es de maîtrise et de tests
res très limitées et utilisés

Ces instruments — que seule la collaboration des ensei-
gnants et des chercheurs permettra de construire en nombre
suffisant — indiquent où l'élève en est (ce qui lui donne sa note)
et où il éprouve des difficultés.

Les remèdes sont multiples et doivent entrer dans notre
pratique pédagogique quotidienne :

- a) Indiquer de façon précise la partie du cours à réétudier.
- b) Le travail par sous-groupes : si un groupe de plus de trois
élèves rencontre la ou les mêmes difficultés, le professeur a
avantage à consacrer une partie du temps de la leçon à ce
groupe, les autres travaillant indépendamment.
- c) Si le groupe compte maximum trois élèves, B. Bloom sug-
gère que chacun d'eux soit intégré dans un groupe hétéro-
gène de deux ou trois. Comme les élèves ne sont plus en
concurrence, mais luttent pour la maîtrise d'une matière,
l'entraide joue.
- d) Dans les écoles comptant plusieurs professeurs d'une même
branche, un élève devrait avoir la faculté de demander une
explication, voire une courte leçon particulière à un autre
professeur que le sien.
- e) Varier la présentation : mettre à la disposition des élèves
des manuels expliquant la notion de différentes façons ;
permettre le recours à l'enseignement programmé, aux
moyens audio-visuels ; varier le niveau d'abstraction.
- f) En général, ne pas fixer à tous le même temps pour les
mêmes choses.

Pareille pratique paie. B. Bloom rapporte une expérience
conduite dans cette ligne où plus de 80 % des élèves ont obtenu
la meilleure note en fin d'année. C'est le triomphe de la péda-
gogie de la courbe en *j*.

Peut-il toujours en être ainsi ? Rien ne permet de l'affirmer.

III. Limite pratique du système.

Même s'il est vrai que tous les élèves de nos écoles sont
capables d'atteindre un haut degré de maîtrise pour autant
qu'on leur laisse un temps suffisant, il est évident qu'au-delà
d'une certaine limite, la durée devient prohibitive.

Comment permettre de passer vingt ans au lieu de six dans
l'enseignement secondaire ! Cette constatation ne doit, toute-
fois, pas remettre le système en cause. Le tout est d'orienter les
élèves en fonction de leurs aptitudes réelles. Le cycle d'obser-
vation et d'orientation, introduit progressivement dans l'ensei-
gnement secondaire de la plupart des pays, en donne la possi-
bilité.

Mais quand, après ce triage, l'étudiant est admis dans une classe, il doit normalement pouvoir apprendre l'essentiel de *tout* ce qu'on y enseigne.

IV. Le système d'enseignement mis en cause.

La pédagogie de la courbe en *j* appelle le nivellement par le haut ; il serait néanmoins chimérique d'imaginer qu'elle permet le nivellement par le... génie.

Nous l'avons vu, moins de cinq pour cent de la population possèdent des aptitudes *exceptionnelles*. Elles doivent être développées au maximum, tant par respect pour la personne que dans l'intérêt de la communauté.

Un nouveau système d'enseignement permet à la fois d'instruire chacun à l'allure convenable et de laisser s'épanouir les surdoués : c'est l'école sans classe, où l'on travaille par groupes homogénéisés selon les aptitudes spécifiques *pour une seule branche*¹.

Dans les conditions actuelles, la classe est un carcan. Pourquoi un élève de six ans qui sait lire en entrant en première primaire ne pourrait-il pas participer aux exercices de lecture de la deuxième année ? Pourquoi tel élève de quatrième de l'enseignement secondaire ne pourrait-il pas suivre un cours de mathématique du niveau de la 3^e ou de la 2^e ?

Dans nos vieilles écoles primaires de campagne, où un seul instituteur se voyait parfois confier les six années, il était commun de voir de tels déplacements. Dans des écoles secondaires qui, actuellement, comptent plusieurs classes de même âge, il est possible de travailler à quatre ou cinq niveaux d'aptitudes différents, au moins dans quelques branches principales. Pareil système fonctionne parfaitement, depuis de nombreuses années, dans des écoles comme la New Trier High School, que nous avons décrite en détail dès 1961².

Prévenons immédiatement l'objection financière. Une école à groupes homogènes spécifiques ne coûte pas plus cher qu'une autre, au contraire. Si l'on ajoute le gain de temps, de talent et la raréfaction des échecs que le système permet, on peut affirmer sans crainte de se tromper que le bénéfice est considérable...

¹ Système à ne pas confondre avec le *streaming* qui consiste à constituer des classes homogènes à l'aide de tests d'aptitudes générales ou, pire encore, en fonction des résultats scolaires globaux. On commet ainsi une double erreur scientifique : dans la grande majorité des cas, un même sujet est inégalement doué pour les diverses branches du programme ; de plus, homogénéiser globalement fait baisser le rendement de l'ensemble. Dans ce système, on aboutit d'ailleurs souvent à une semi-ségrégation sociale. Tous ces inconvénients ont été si bien mis en évidence que le *streaming* est interdit par la loi en Suède.

² *L'individualisation de l'enseignement dans une école multilatérale*, in Education, mai 1961.

tudiant est admis dans une
r apprendre l'essentiel de

en cause.

appelé le nivellement par
que d'imaginer qu'elle per-

pour cent de la population
elles. Elles doivent être dé-
spect pour la personne que

ement permet à la fois d'in-
et de laisser s'épanouir les
, où l'on travaille par grou-
des spécifiques pour une

la classe est un carcan.
t lire en entrant en première
r aux exercices de lecture
tel élève de quatrième de
ait-il pas suivre un cours de
de la 2^e ?

aires de campagne, où un
nfier les six années, il était
nts. Dans des écoles secon-
plusieurs classes de même
atre ou cinq niveaux d'apti-
lques branches principales.
ent, depuis de nombreuses
New Trier High School, que
61².

ection financière. Une école
ne coûte pas plus cher
ajoute le gain de temps, de
que le système permet, on
omper que le bénéfice est

streaming qui consiste à constituer des
générales ou, pire encore, en fonction
une double erreur scientifique : dans
inégalement doué pour les diverses
globalement fait baisser le rendement
leurs souvent à une semi-ségrégation
mis en évidence que le streaming est

une école multilatérale, in Education,

CONCLUSIONS ET RECOMMANDATIONS

Les conclusions partielles se nous avançons dans cette étude pourtant loin d'avoir épuisé le apporté des solutions définitives.

Sortir des routines irréfléchies les grandes théories pédagogiques de l'éducation, découvre tant de remplir que l'on risque d'être enragement !

Au risque de laisser nos lecteurs pourtant pas voulu voiler la difficulté croire à l'existence d'une doctrine conduire à une évaluation par des moyens. Vraisemblablement, et la mesure rigoureuse des traits humains jours impossible : le sort nous guérit âmes !

Que l'évaluation continue s'intègre au processus d'enseignement et d'apprentissage, n'en supprime pas la dimension d'une docimastique, c'est-à-dire de notation, autrement dit l'évaluation systématique du rendement.

Par ailleurs, l'exigence d'évaluation peut regretter pour des raisons de nature scolaire : elle correspond à notre civilisation.

Peut-être en sera-t-il un jour qui continue à s'intellectualiser, et nous ne voyons toutefois pas complètement aux examens et à la notation.

Notre position docimologique dans son désir de concilier les deux nouvelles, de nature surtonique et fondamentale de l'analyse.

Pour nous limiter à un seul aspect nous ne pouvons concevoir que les prises venant en fin de chaque année soient prises seulement en fonction de choix multiple se prêtant à la notation. L'introduction de ce type d'épreuve pour le contrôle objectif et apprivoisé il ne suffit pas. L'évaluation du travail et la rencontre ultime du maître-élève les nuances que l'approche quantitative

Les conclusions partielles se sont imposées à mesure que nous avançons dans cette étude déjà longue. Nous sommes pourtant loin d'avoir épuisé le sujet et, encore plus, d'avoir apporté des solutions définitives.

Sortir des routines irréfléchies et ambitionner de traduire les grandes théories pédagogiques dans la pratique quotidienne de l'éducation, découvre tant d'écueils, tant de conditions à remplir que l'on risque d'être envahi par le doute et le découragement !

Au risque de laisser nos lecteurs insatisfaits, nous n'avons pourtant pas voulu voiler la difficulté, et moins encore laisser croire à l'existence d'une docimologie achevée, capable de conduire à une évaluation parfaite, si on lui en donne les moyens. Vraisemblablement, et heureusement sans doute, la mesure rigoureuse des traits humains les plus fins restera toujours impossible : le sort nous garde de la machine à peser les âmes !

Que l'évaluation continue s'insère fonctionnellement dans le processus d'enseignement et d'apprentissage dont elle est indissociable, n'en supprime pas, pour autant, la nécessité d'une docimastique, c'est-à-dire d'une technique d'examen et de notation, autrement dit encore, d'une technique d'observation systématique du rendement immédiat ou différé.

Par ailleurs, l'exigence d'évaluation comparée, que l'éducateur peut regretter pour des raisons idéales, dépasse le domaine scolaire : elle correspond à un caractère fondamental de notre civilisation.

Peut-être en sera-t-il un jour autrement. Si notre culture continue à s'intellectualiser, et donc à croître en complexité, nous ne voyons toutefois pas comment elle pourrait renoncer complètement aux examens et aux concours.

Notre position docimologique est résolument éclectique dans son désir de concilier les avantages certains de techniques nouvelles, de nature surtout statistique, et l'apport indéniable et fondamental de l'analyse qualitative.

Pour nous limiter à un seul exemple de cet éclectisme, nous ne pouvons concevoir que les décisions de passage intervenant en fin de chaque année d'études universitaires soient prises seulement en fonction des résultats obtenus à un test à choix multiple se prêtant à la notation automatique. Assurément, l'introduction de ce type d'épreuves est hautement souhaitable pour le contrôle objectif et approfondi des connaissances, mais il ne suffit pas. L'évaluation du travail de l'année doit intervenir et la rencontre ultime du maître et de son élève apporter toutes les nuances que l'approche quantitative a laissées dans l'ombre.

De même que l'introduction des machines dans l'industrie a permis à l'intelligence de prendre le pas sur la force musculaire et la routine avilissante, de même le contrôle automatique de la connaissance des faits, des méthodes et des techniques rend possible un examen final réellement centré sur les processus mentaux et les traits de personnalité les plus nobles. Quelle importance respective accordera-t-on aux trois types d'évaluation ? La décision appartient, dans ce cas, à la communauté académique, éclairée par les recherches docimologiques.

Un certain nombre de *recommandations* paraissent s'imposer.

En raison même de nos ignorances et de la grande complexité de l'évaluation, il est d'abord urgent qu'un *réseau de centres de recherche en éducation* couvre le pays entier et soit mis au service du système scolaire. Les fonctions de ces centres sont apparues au cours de notre analyse : recherches portant sur les problèmes fondamentaux et développement des méthodes, des techniques et des instruments nécessaires. Ces centres ne doivent pas être greffés en corps étrangers dans l'ensemble de nos institutions éducatives, mais bien travailler en étroite collaboration avec elles.

Plus particulièrement, *chaque université* devrait créer en son sein un *bureau d'études des problèmes d'enseignement et d'examens*. On ne peut concevoir que l'institution scientifique par excellence exclue de la rigueur de son analyse une de ses deux raisons d'être : l'éducation.

La *définition des objectifs* généraux et spéciaux, fondement de tout l'édifice de l'enseignement et de l'évaluation, requiert le travail en commun d'un grand nombre de chercheurs et de praticiens de toutes les disciplines. *Des commissions spécialisées* permettront la coopération nécessaire et la communication, — dans les deux sens, — entre la base et le sommet de la hiérarchie pédagogique.

La réforme des examens et des modes d'évaluation entraînera sans doute la disparition définitive du calcul du pourcentage, presque universellement répandu dans nos écoles jusqu'à ces temps derniers. Quelles que soient ses faiblesses, l'échelle des pourcentages offrait cependant l'avantage d'être familière à tous et chacun avait au moins l'illusion de la comprendre. Actuellement, les échelles les plus diverses sont essayées dans nos écoles. *Une unification du système de notation* et *une campagne d'information* sont nécessaires dans le plus bref délai, faute de quoi la communication deviendra

malaisée, sinon impossible, entre différentes écoles, entre les écoles avant de prendre une décision d'experts de la docimologie et, plus en sciences humaines devraient

Un vaste effort d'information des enseignants appelés à changer des d'évaluation. Le moyen le plus l'*organisation d'expériences locales* des examens, la notation et la seraient, par exemple, incorporé à l'introduction de l'enseignement, et dans la réforme de la formation, toutes les écoles y participeraient.

Des commissions de modération seront petit à petit constituées.

Une action trop hâtive risquerait d'être éminemment louable. On ne doit pas décevoir ou d'égaler les enseignants sous l'autorité de docimologistes.

Œuvre humaine par excellence, la mesure de la qualité que dans la mesure où elle constitue le fondement des idées nouvelles et des méthodes mêmes.

des machines dans l'industrie
prendre le pas sur la force musculaire
même le contrôle automatique
des méthodes et des techniques
réellement centré sur les pro-
personnalité les plus nobles.
accordera-t-on aux trois types
partient, dans ce cas, à la com-
par les recherches docimolo-

mandations paraissent s'impo-

norances et de la grande com-
l'abord urgent qu'un *réseau de*
tion couvre le pays entier et soit
laire. Les fonctions de ces cen-
notre analyse : recherches por-
nementaux et développement des
es instruments nécessaires. Ces
effés en corps étrangers dans
ducatives, mais bien travailler en

que *université* devrait créer en
s problèmes d'enseignement et
voir que l'institution scientifique
leur de son analyse une de ses
n.

généraux et spéciaux, fondement
ent et de l'évaluation, requiert le
ombre de chercheurs et de pra-
Des commissions spécialisées
essaire et la communication, —
base et le sommet de la hiéar-

des modes d'évaluation entraî-
é définitive du calcul du pourcen-
épandu dans nos écoles jus-
es que soient ses faiblesses,
ait cependant l'avantage d'être
au moins l'illusion de la com-
elles les plus diverses sont
unification du système de nota-
tion sont nécessaires dans le
la communication deviendra

malaisée, sinon impossible, entre l'école et la famille, entre les
différentes écoles, entre les écoles et les autorités. Pourtant,
avant de prendre une décision dont l'importance est patente, les
experts de la docimologie et, plus généralement, de la mesure
en sciences humaines devraient être entendus.

Un vaste effort d'information sera aussi entrepris auprès
des enseignants appelés à changer profondément leurs habitu-
des d'évaluation. Le moyen le plus efficace semble résider dans
l'organisation d'expériences locales portant sur la préparation
des examens, la notation et la modération. Ces expériences
seraient, par exemple, incorporées dans les travaux nécessaires
à l'introduction de l'enseignement secondaire rénové, en Belgi-
que, et dans la réforme de la formation des maîtres ; progressi-
vement, toutes les écoles y participeraient.

Des commissions de modération locales et régionales
seront petit à petit constituées.

Une action trop hâtive risque de mettre en péril une réno-
vation éminemment louable. On se gardera, en particulier, de
décevoir ou d'égarer les enseignants en tentant de les placer
sous l'autorité de docimologistes improvisés.

Œuvre humaine par excellence, l'éducation ne gagnera en
qualité que dans la mesure où les éducateurs accepteront pro-
fondément les idées nouvelles et tenteront de se dépasser eux-
mêmes.

ANN

ANNEXES

**ETUDE COMPAREE D'UNE
PRESENTEE SELON LA
ET SELON
PAR QUESTIONS**

*Exposé de la question*¹ (Mét)

- 1° a) Donner la liste des causes
b) Décrire les lésions anatomiques
l'hypoglycémie.

Pour permettre aux correcteurs
mité dans l'appréciation, la
l'ensemble des correcteurs :

- A. Valeur égale pour les deux
B. Pour obtenir une note de 10
qué dans la liste les causes

- 1) Excès d'insuline ;
- 2) Tumeur fonctionnelle de
pancréas ;
- 3) Influence endocrinienne
hypophyse et des surrénales ;
- 4) Maladie du foie empêchant
gène (nécrose aiguë) ou
maladie de Von Gierke).

- 2° A quoi reconnaît-on la différence
- une seule crise aiguë d'hypoglycémie
- une hypoglycémie chronique

¹ MOORE, Robert A., *Methods of Medical Education*, janvier 1954, vol. 29, n° 1.
Traduction empruntée au rapport français
o.c.

**ETUDE COMPAREE D'UNE QUESTION D'EXAMEN
PRESENTEE SELON LA METHODE TRADITIONNELLE
ET SELON LA METHODE
PAR QUESTIONS A CHOIX MULTIPLE.**

*Exposé de la question*¹ (Méthode traditionnelle).

- 1° a) Donner la liste des causes d'hypoglycémie ;
b) Décrire les lésions anatomo-pathologiques résultant de l'hypoglycémie.

Pour permettre aux correcteurs un certain degré d'uniformité dans l'appréciation, la grille suivante a été établie par l'ensemble des correcteurs :

- A. Valeur égale pour les deux parties de la question ;
B. Pour obtenir une note de 75/100, le candidat doit avoir indiqué dans la liste les causes :

- 1) Excès d'insuline ;
- 2) Tumeur fonctionnelle des îlots de Langerhans ;
- 3) Influence endocrinienne par hypofonctionnement de l'hypophyse et des surrénales ;
- 4) Maladie du foie empêchant la mise en réserve du glycogène (nécrose aiguë) ou la libération du glycogène (maladie de Von Gierke).

- 2° A quoi reconnaît-on la différence des lésions causées par une seule crise aiguë d'hypoglycémie de celles causées par une hypoglycémie chronique ?

¹ MOORE, Robert A., *Methods of Examining Students in Medicine*, in « Journal of Medical Education », janvier 1954, vol. 29, n° 1.
Traduction empruntée au rapport français sur la réforme des examens de médecine, o.c.

3° Décrire les modifications visibles au niveau du cerveau.

C. Pour obtenir une note allant de 75 à 90/100, le candidat doit avoir indiqué :

- soit une liste montrant qu'il a compris le mécanisme d'action de chacune des causes,
- soit une liste comprenant d'autres causes, mais en indiquant qu'il comprend que les mécanismes d'homéostasie ont habituellement tendance à corriger l'hypoglycémie, dans les cas tels que :

- 1) Ingesta diminués ;
- 2) Absorption perturbée ;
- 3) Métabolisme augmenté comme dans l'hyperthyroïdisme ;
- 4) Infection grave ;
- 5) Surmenage physique ;
- 6) Glycosurie rénale.

D. Pour obtenir une note supérieure à 90/100, le candidat doit avoir indiqué :

- une liste logique des quatre causes majeures et des six causes mineures, en plus d'un type idiopathique,
- une différenciation de l'hypoglycémie aiguë et chronique par une description détaillée des lésions cérébrales.

Exposé de la question (Méthode par questions à choix multiple).

Voici comment une même question pourrait être présentée selon la nouvelle méthode.

1° On peut déterminer si le candidat est capable de reconnaître les quatre causes les plus importantes et les différencier des moins importantes dans la question suivante :

INSTRUCTIONS. Quatre des cinq phrases numérotées sont communes à l'un des trois troubles fonctionnels de la liste alphabétique (A.B.C.).

Indiquez celle qui est l'exception et le trouble fonctionnel commun aux quatre phrases restantes.

- | | |
|---|---|
| 1. Excès d'insuline | A. Hypoglycémie cliniquement décelable |
| 2. Tumeur fonctionnelle des îlots de Langerhans | B. Hyperglycémie cliniquement décelable |
| 3. Glycosurie rénale | C. Glycosurie cliniquement décelable |
| 4. Hypofonctionnement hypophysaire | |
| 5. Maladie de Von Gierke | |

Si le candidat répond correctement, 1, 2, 4 et 5 peuvent produire une hypoglycémie décelable, que ce n'est pas la raison des quatre cas par rapport à l'hyperglycémie, ni à la possession de connaissances prouvées.

2° Si l'on veut savoir ce que les hormones qui commandent l'équilibre du glucose et de l'insuline, on peut poser la question suivante :

INSTRUCTIONS. Chacune des phrases ci-dessous est composée de deux parties : la première énonce un fait ; la seconde en indique la cause.

Indiquez pour chacune des phrases A, B, C, D ou E, selon que :

- A. L'exposé du fait et de la cause est une relation de cause à effet.
- B. L'exposé du fait et de la cause est une relation de conséquence à cause.
- C. L'exposé du fait est une relation de cause à effet.
- D. L'exposé du fait est une relation de conséquence à effet ou un principe accepté.
- E. L'exposé du fait et la cause sont indépendants.

1. Le taux du sucre sanguin est élevé après une hépatectomie parce que le foie est la source principale de glucose.

(E) 2. Le diagnostic anatomique de la maladie de Langerhans est hyperglycémie parce que la maladie était hyperadénomateuse des îlots de Langerhans.

(C) 3. Les malades atteints de diabète ont de meilleures chances d'avoir une hypoglycémie parce qu'ils ont un hyperinsulinisme.

(C) 4. Les malades atteints de diabète sentent une diminution de l'appétit parce que, dans cette maladie, le glucose est mis en réserve par le foie.

(E) 5. Une hypoglycémie de longue durée n'est pas suivie de séquestration de glucose dans les cellules adipeuses produites par le foie.

bles au niveau du cerveau.

de 75 à 90/100, le candidat

u'il a compris le mécanisme
causes,

nt d'autres causes, mais en
l que les mécanismes d'ho-
ment tendance à corriger
cas tels que :

é comme dans l'hyperthyroï-

ure à 90/100, le candidat

e causes majeures et des
s d'un type idiopathique,
oglycémie aiguë et chroni-
llée des lésions cérébrales.

questions à choix multiple).

tion pourrait être présentée

est capable de reconnaître
portantes et les différencier
uestion suivante :

q phrases numérotées sont
troubles fonctionnels de la

on et le trouble fonctionnel
stantes.

- a. Hypoglycémie clinique-
ment décelable
- b. Hyperglycémie clinique-
ment décelable
- c. Glycosurie cliniquement
décelable

Si le candidat répond correctement, il montre qu'il sait que 1, 2, 4 et 5 peuvent produire une hypoglycémie cliniquement décelable, que ce n'est pas le cas de 3, et qu'aucune combinaison des quatre cas parmi les cinq ne peut être associée ni à l'hyperglycémie, ni à la glycosurie. En d'autres termes, la possession de connaissances positives et négatives est prouvée.

- 2° Si l'on veut savoir ce que sait le candidat sur les phénomènes qui commandent l'équilibre du niveau du sucre sanguin et de l'insuline, on peut poser la question suivante :

INSTRUCTIONS. Chacune des phrases suivantes est composée de deux parties : l'exposé d'un fait et la raison de ce fait.

Indiquez pour chacune des phrases numérotées la lettre A, B, C, D ou E, selon que :

- A. L'exposé du fait et sa raison sont vrais et ont une relation de cause à effet.
 - B. L'exposé du fait et sa raison sont vrais, mais n'ont pas de relation de cause à effet.
 - C. L'exposé du fait est vrai, mais la raison est fausse.
 - D. L'exposé du fait est faux, mais la raison est un fait ou un principe accepté.
 - E. L'exposé du fait et la raison sont faux.
1. Le taux du sucre sanguin tombe brutalement après hépatectomie parce que le glycogène contenu dans le foie est la source principale du sucre sanguin.
(A)
 2. Le diagnostic anatomo-pathologique d'un adénome des îlots de Langerhans du pancréas implique que la maladie était hyperinsulinique parce que tous les adénomes des îlots sont fonctionnels et secrètent de l'insuline.
(E)
 3. Les malades atteints d'hyperthyroïdisme ont toutes chances d'avoir une hypoglycémie parce qu'il existe un hyperinsulinisme associé.
(C)
 4. Les malades atteints de la maladie de Von Gierke présentent une diminution du taux du sucre sanguin parce que, dans cette maladie, le glycogène n'est pas mis en réserve par le foie.
(C)
 5. Une hypoglycémie durant depuis plusieurs mois n'est pas suivie de séquelles parce que les modifications cellulaires produites par l'hypoglycémie sont réversibles.
(E)

Si le candidat répond correctement à cette série, cela montre qu'il connaît :

- a) que le glycogène hépatique est la source principale permettant le maintien du taux du sucre sanguin ;
- b) que toutes les tumeurs des îlots de Langerhans ne sont pas fonctionnelles ;
- c) que les malades atteints d'hyperthyroïdie ont une hypoglycémie, mais que la cause n'est pas un hyperinsulinisme associé ;
- d) que les malades atteints de la maladie de Von Gierke ont une hypoglycémie, mais qu'elle n'est pas due au fait qu'il n'y a pas de glycogène dans le foie ;
- e) qu'il y a des séquelles après hypoglycémie chronique et que les modifications cellulaires résultantes ne sont pas réversibles.

3° On peut déterminer si le candidat a quelques notions simples concernant les modifications au niveau du cerveau dans la question suivante :

INSTRUCTIONS. Chacun des exposés incomplets (numérotés) est suivi de cinq compléments au choix. Indiquez, dans chaque cas, le complément qui convient le mieux.

1. Les modifications anatomiques secondaires à une hypoglycémie chronique sont mises en évidence le plus souvent dans :
 - a) la rate
 - b) le rein
 - c) les surrénales
 - d) le cerveau
 - e) la thyroïde
2. Parmi les modifications provoquées par l'hypoglycémie chronique au niveau du cerveau, les plus importantes sont sur :
 - a) les neurones
 - b) les astrocytes
 - c) les cellules de l'épendyme
 - d) les cellules oligodendrogliques
 - e) les cellules microgliales

3. Parmi les altérations provoquées par l'hypoglycémie :
 - a) le déplacement du noyau
 - b) la disparition de la chromatine
 - c) l'altération de la structure des mitochondries
 - d) l'éclatement du noyau
 - e) la fusion de mitochondries

4. Parmi les modifications du cerveau par l'hypoglycémie chronique, la plus caractéristique est :
 - a) l'hydrocéphalie interne
 - b) l'épaississement de la paroi des vaisseaux
 - c) la destruction des cellules de Nissl
 - d) la prolifération des cellules gliales
 - e) la prolifération des cellules de Schwann

Si le candidat répond correctement à cette question, cela montre qu'il sait que les altérations du cerveau par l'hypoglycémie chronique atteignent les neurones et que l'hypoglycémie aiguë provoque une destruction des cellules de Nissl des neurones et que la destruction de cellules de Nissl est irréversible.

4° On peut déterminer si le candidat a quelques notions simples concernant les troubles, dans la question suivante :

INSTRUCTIONS. Dans la question suivante, cinq mécanismes différents sont proposés. Indiquez, dans chaque cas, le mécanisme qui convient le mieux.

Inscrivez la lettre appropriée à côté de chaque question, numérotées en associant la lettre à la question.

- A. Augmentation de l'activité de l'enzyme adénylate cyclase.
 - B. Hyperinsulinisme.
 - C. Mise en réserve d'un produit.
 - D. Absence de mise en réserve d'un produit.
 - E. Hypofonctionnement des cellules de Nissl.
- (B) 1. Adénomes des îlots de Langerhans.
- (A) 2. Exercice physique.
- (A) 3. Hyperthyroïdisme.

ement à cette série, cela mon-

est la source principale per-

du sucre sanguin ;

les îlots de Langerhans ne sont

d'hyperthyroïdie ont une hypo-

use n'est pas un hyperinsuli-

la maladie de Von Gierke ont

qu'elle n'est pas due au fait

dans le foie ;

ès hypoglycémie chronique et

laïres résultantes ne sont pas

didat a quelques notions sim-

itions au niveau du cerveau

exposés incomplets (numéro-

pléments au choix. Indiquez,

ément qui convient le mieux.

miques secondaires à une hy-

ont mises en évidence le plus

provoquées par l'hypoglycémie

du cerveau, les plus impor-

yme

drogiques

es

3. Parmi les altérations des cellules nerveuses provoquées par l'hypoglycémie aiguë, la plus évidente est :

- a) le déplacement du noyau
- b) la disparition de la paroi cellulaire
- c) l'altération de la substance de Nissl
- d) l'éclatement du noyau de la cellule
- e) la fusion de mitochondries.

4. Parmi les modifications provoquées au niveau du cerveau par l'hypoglycémie chronique, la plus significative est :

- a) l'hydrocéphalie interne
- b) l'épaississement fibreux de l'arachnoïde
- c) la destruction des neurones
- d) la prolifération des astrocytes
- e) la prolifération des cellules de l'épendyme

Si le candidat répond correctement à cette série, il montre qu'il sait que les altérations principales de l'hypoglycémie chronique atteignent les neurones du cerveau, que l'hypoglycémie aiguë provoque une altération de la substance de Nissl des neurones et que l'hypoglycémie chronique provoque la destruction de cellules nerveuses.

4° On peut déterminer si le candidat comprend le mécanisme des troubles, dans la question suivante :

INSTRUCTIONS. Dans la liste alphabétique sont indiqués cinq mécanismes différents pouvant conduire à l'hypoglycémie.

Inscrire la lettre appropriée après chacun des exposés numérotés en associant au trouble le mécanisme responsable :

- A. Augmentation de l'activité métabolique.
- B. Hyperinsulinisme.
- C. Mise en réserve d'un glycogène anormal dans le foie.
- D. Absence de mise en réserve de glycogène dans le foie.
- E. Hypofonctionnement de l'hypophyse ou des surrénales.

- (B) 1. Adénomes des îlots de Langerhans.
- (A) 2. Exercice physique violent.
- (A) 3. Hyperthyroïdisme.

- (E) 4. Maladie de Simmonds.
- (C) 5. Maladie de Von Gierke.
- (D) 6. Hépatite épidémique.
- (C) 7. Maladie d'Addison.

Si le candidat répond correctement à cette série, cela montre qu'il comprend les bases du métabolisme des glucides et connaît les facteurs influençant ce métabolisme.

Récapitulation.

Ainsi, au moyen de ces dix-sept « questions objectives », nous avons mis en évidence toutes les connaissances requises pour pouvoir donner une note supérieure à 90/100, à savoir :

1. Liste des quatre causes principales.
2. Reconnaître la différence entre les effets de l'hypoglycémie chronique et aiguë.
3. Décrire les altérations au niveau du cerveau.
4. Compréhension des mécanismes.
5. Liste des causes mineures.
6. Différencier les lésions de l'hypoglycémie aiguë et chronique.

COMPOSITION DE
 ANGLETERRE — G
 OF EDUC
 FIN DU SECONDAIRE

EPREUVE I. Deux questions —

1. Choisissez un des sujets suivants :
 - a. Une nuit brumeuse.
 - b. Un marchand achète et vend d'occasion. Décrivez les circonstances.
 - c. Quelle serait votre politique en matière de radio ou de télévision ?
 - d. Une grande foule se déplace. Décrivez les circonstances.
 - e. Plaisir de la photographie ou du cyclisme.
 - f. Pensez-vous que garçons et filles ont des carrières ?
 - g. Quelles sont vos réactions à la réalisation de l'exploration spatiale ?

2. Choisissez un des sujets suivants :
 - a. Décrivez un entraînement intensif dans un sport de votre choix.
 - b. Après avoir visité une école, écrivez un article pour vos compagnons sur les possibilités d'avenir qu'elle offre.
 - c. Décrivez clairement un appareil électroménager. Expliquez comment il fonctionne. (sèche-cheveux, un aspirateur, un mixer).
 - d. Dans votre ville, on veut améliorer la circulation. Écrivez une lettre à un journal local pour proposer un projet.

**COMPOSITION DE LANGUE MATERNELLE
ANGLETERRE — GENERAL CERTIFICATE
OF EDUCATION 1967
FIN DU SECONDAIRE — NIVEAU ORDINAIRE.**

EPREUVE I. Deux questions — 1½ h.

1. Choisissez un des sujets suivants. Consacrez-y environ 1 h.
 - a. Une nuit brumeuse.
 - b. Un marchand achète et revend, le même jour, un objet d'occasion. Décrivez les deux scènes.
 - c. Quelle serait votre politique si vous dirigiez les programmes de radio ou de télévision ?
 - d. Une grande foule se disperse. Décrivez la scène.
 - e. Plaisir de la photographie *ou* du dessin *ou* de la danse *ou* du cyclisme.
 - f. Pensez-vous que garçons et filles ont les mêmes chances de carrière ?
 - g. Quelles sont vos réactions devant les progrès et les réalisations de l'exploration spatiale ?

2. Choisissez un des sujets suivants. Consacrez-y environ ½ h.
 - a. Décrivez un entraînement destiné à améliorer vos performances dans un sport de votre choix.
 - b. Après avoir visité une entreprise, faites un rapport à vos compagnons sur les conditions de travail et les perspectives d'avenir qu'elle offre.
 - c. Décrivez clairement *un* des appareils suivants et expliquez comment il fonctionne : un « walkie talkie », un sèche-cheveux, un aspirateur, un mélangeur-batteur (mixer).
 - d. Dans votre ville, on veut créer un centre commercial où la circulation sera interdite aux véhicules. Ecrivez une lettre à un journal local pour exposer vos vues sur le projet.

EPREUVE II. Quatre questions — 1³/₄ h.

1. Résumez le passage suivant en bonne prose continue et en 110 mots maximum. A la fin, indiquez combien de mots vous avez utilisés. Le passage compte 314 mots.

Suit un texte sur la confiance exagérée dans la science et la technologie.

2. Lisez le passage suivant ; ensuite, répondez aux questions.

(Le texte décrit deux grands types de promeneurs — ceux qui suivent un guide et ceux qui partent à l'aventure — et souligne l'intérêt des excursions géologiques.)

- a. Expliquez avec vos propres mots la différence entre les deux types de promeneurs.
- b. Quel conseil l'auteur donne-t-il à propos des grottes ? Formulez-le avec vos propres mots.
- c. Expliquez brièvement pourquoi l'auteur croit que la géologie est : un *hobby* amusant — un *hobby* instructif.
- d. L'auteur écrit qu'il utilise l'expression « creuser un fossé » au sens littéral. Expliquez pourquoi le sens est ici littéral.
- e. Expliquez les expressions suivantes : précautions prescrites ; être parfaitement conscient de la nature de ses actes.
- f. Choisissez quatre des mots suivants. Remplacez-les par des synonymes ou des périphrases qui pourraient être utilisées dans le texte sans en changer le sens (...).
3. Répondez, au choix, à une des deux questions suivantes :
- a. Choisissez trois mots parmi les suivants. Construisez des phrases (6 en tout) montrant que ces mots peuvent être employés dans deux sens différents (...).
- b. Définissez en une phrase trois des mots suivants : monopole - interlude - préface - microscope - antidote.
4. Répondez au choix à une des deux questions suivantes :
- a. Expliquez clairement, mais brièvement, la différence de sens entre chaque paire de phrases (porte sur *could-should* ; *can-may* ; *will-shall* ; *might-must*).
- b. Réécrivez correctement le passage suivant, en respectant toutes les idées. Vous pouvez changer l'expression, l'ordre des mots et des idées, l'orthographe et la ponctuation.
(Suit un texte défectueux d'une centaine de mots).

« Par exemple si
à travers une ville
dans son chemin
gés, entraînant
contourner le bât
et la rendant aus
déviation de cette
causés mais le p
spécialement si l
la distance ajout
camion portant c
camion consomm
plus de temps po
enlevés de sa vie

1³/₄ h.

en bonne prose continue et en indiquez combien de mots vous avez écrits (314 mots).

« *La science exagérée dans la science* »

Ensuite, répondez aux questions.

« *Les types de promeneurs — et ceux qui partent à l'aventure des excursions géologiques.* »

Indiquez la différence entre les

« *grottes* » et « *grottes* » ?

« *grottes* » — un *hobby* instructif.

« *grottes* » — pourquoi le sens est ici

« *grottes* » : précautions prises par un géologue conscient de la nature de ses

« *grottes* ». Remplacez-les par des phrases qui pourraient être utilisées en changeant le sens (...).

« *grottes* » : deux questions suivantes :

« *grottes* » : les suivantes. Construisez des phrases qui peuvent être utilisées en changeant le sens (...).

« *grottes* » : des mots suivants : microscope - antidote.

« *grottes* » : deux questions suivantes :

« *grottes* » : brièvement, la différence de sens entre les phrases (porte sur *could-might-must*).

« *grottes* » : message suivant, en respectant le message, changez l'expression, l'orthographe et la ponctuation.

« *grottes* » : (une centaine de mots).

« Par exemple si on projetait de faire passer une route à travers une ville mais qu'une maison historique était dans son chemin alors les plans devraient être changés, entraînant des dépenses considérables, pour contourner le bâtiment créant un virage dans la route et la rendant aussi dangereuse pour les autos. Par la déviation de cette route non seulement des frais sont causés mais le prix de différentes choses augmente spécialement si la déviation est grande parce que si la distance ajoutée est disons dix kilomètres et qu'un camion portant certains articles parcourt la route, le camion consommerait alors plus d'essence prendrait plus de temps pour arriver et dix kilomètres seraient enlevés de sa vie. »

CONSIGNES POUR LA CORRECTION.

EPREUVE I. — Maximum 50 points.

Le schéma de notation qui figure ci-dessous ne constitue qu'un guide préliminaire. Des additions et des amendements pourront être apportés lors de la réunion des examinateurs qui sera convoquée après une première lecture des travaux.

Question 1 (maximum 35 points).

On attend un minimum de 400 mots, mais les compositions ne doivent pas être principalement notées en fonction de la longueur. Tenir compte du sujet choisi et de la façon dont il est traité. Même si elle est courte, une composition où l'argumentation est serrée et où l'expression est bonne doit obtenir plus de points qu'une longue narration informe.

L'examineur devrait avoir une idée claire de ce qu'est une composition recevant tout juste la note de passage (16 points). Pareille composition doit contenir des idées raisonnables, mais non très originales. L'expression doit être claire, mais sans distinction particulière. On ne devrait rencontrer, dans le travail, que quelques erreurs mécaniques. Les candidats qui ont dépassé ce niveau général doivent être récompensés et ceux qui ont travaillé en dessous de ce niveau doivent être pénalisés.

On trouvera dans le document annexé des notes détaillées sur les qualités à observer dans les compositions. On compte toutefois que les examinateurs noteront en fonction de leur impression générale, n'alloueront donc pas une proportion fixe des points pour les différents aspects. Si le sujet se prête à la controverse, les idées et leur enchaînement peuvent être plus importants que dans une composition descriptive où le vocabulaire pourrait prendre une plus grande place. Nous nous fions au jugement des examinateurs.

En réservant 35 points à cette première question, on a voulu marquer son importance majeure dans l'ensemble des examens. Beaucoup de candidats sont médiocres. Toutefois, si l'examineur ne disperse pas largement ses notes, cette première question ne pèsera pas d'un poids suffisant dans l'ensemble des résultats.

Groupe A (29-35 points).

La composition est de qualité exceptionnelle.

Groupe B (22-28 points).

La qualité du travail est au-dessus de la moyenne.

Groupe C (14-21 points).
Le travail est de qualité m

Groupe D (7-13 points).
Le travail est d'un niveau
comme satisfaisant.

Groupe E (0-6 points).
Le candidat est incapable
cohérence.

Cette épreuve est un exa
l'épreuve 1, nous examinons la
ses opinions, ses expériences
et ses intérêts. Il ne s'agit ni
rales, ni d'une évaluation de
jeune fille décrit clairement
obtenir une note favorable si
nents, même si l'examineur
tion aborde le sujet de façon
si, derrière la masse de pro
l'impression que se trouvent
attitudes morales élevées, il n
aspect : seul, ce qui est écrit
la clarté de l'expression et la

Question 1 (Maximum 35

- (a) Une narration ou une de
une narration, la nuit brum
- (b) En gros, les deux scènes
penser la vivacité de narra
et le contraste.
- (c) On attend une définition cl
points doivent être déve
choisi la radio ou la télév
deux.
- (d) Le sujet du travail est la
une courte introduction.
- (e) Une réponse cohérente, d
plus que de longs errement
- (f) Ce sujet n'est pas facile,
bonne ordonnance des a
choisis.
- (g) Un traitement purement na
mais quelques exemples d
nécessaires pour expliquer

A CORRECTION.

Figure ci-dessous ne constitue
ditions et des amendements
réunion des examinateurs qui
re lecture des travaux.

mots, mais les compositions
nt notées en fonction de la
choisi et de la façon dont il
une composition où l'argu-
ssion est bonne doit obtenir
on informe.

ne idée claire de ce qu'est
ste la note de passage (16
contenir des idées raisonna-
ression doit être claire, mais
devrait rencontrer, dans le
iques. Les candidats qui ont
être récompensés et ceux
iveau doivent être pénalisés.

annexé des notes détaillées
s compositions. On compte
teront en fonction de leur
onc pas une proportion fixe
ts. Si le sujet se prête à la
ainement peuvent être plus
ion descriptive où le voca-
ande place. Nous nous fions

première question, on a voulu
ns l'ensemble des examens.
res. Toutefois, si l'examina-
ses notes, cette première
suffisant dans l'ensemble

exceptionnelle.

sus de la moyenne.

Groupe C (14-21 points).

Le travail est de qualité moyenne.

Groupe D (7-13 points).

Le travail est d'un niveau inférieur à ce que l'on considère
comme satisfaisant.

Groupe E (0-6 points).

Le candidat est incapable de présenter ses idées avec
cohérence.

Cette épreuve est un examen de langue maternelle. Dans
l'épreuve 1, nous examinons la capacité du candidat à exprimer
ses opinions, ses expériences, ses impressions, ses sentiments
et ses intérêts. Il ne s'agit ni d'un test de connaissances géné-
rales, ni d'une évaluation des aptitudes du candidat. Si une
jeune fille décrit clairement une scène en anglais, elle doit
obtenir une note favorable si les matériaux utilisés sont perti-
nents, même si l'examinateur pense que la jeune fille en ques-
tion aborde le sujet de façon trop sentimentale. Par ailleurs,
si, derrière la masse de prose incohérente, l'examinateur a
l'impression que se trouvent des émotions profondes et des
attitudes morales élevées, il n'a pas à s'occuper de ce dernier
aspect : seul, ce qui est écrit compte. Pour réussir l'épreuve,
la clarté de l'expression et la précision du style sont essentiels.

Question 1 (Maximum 35 points)

- (a) Une narration ou une description sont acceptables. Dans
une narration, la nuit brumeuse doit jouer un rôle essentiel.
- (b) En gros, les deux scènes doivent être équilibrées. Récom-
penser la vivacité de narration, de dialogue, de description
et le contraste.
- (c) On attend une définition claire de la politique dont plusieurs
points doivent être développés. Le candidat doit avoir
choisi la radio ou la télévision ; il ne peut avoir réuni les
deux.
- (d) Le sujet du travail est la dispersion ; autoriser néanmoins
une courte introduction.
- (e) Une réponse cohérente, claire, de longueur modérée fait
plus que de longs errements et des répétitions.
- (f) Ce sujet n'est pas facile, récompenser généreusement la
bonne ordonnance des arguments et les exemples bien
choisis.
- (g) Un traitement purement narratif ne peut pas être accepté,
mais quelques exemples de progrès réalisés peuvent être
nécessaires pour expliquer les réactions.

Question 2 (Maximum 15 points).

- (a) On attend une description claire et logique.
- (b) Le rapport doit traiter les trois aspects de la question. On insiste surtout sur l'information rapportée ; on n'exigera donc pas une forme de rapport particulier.
- (c) On exige à la fois une description de l'objet et une explication de son fonctionnement. Ne punissez pas sévèrement les erreurs matérielles ; notez simplement la clarté de l'expression.
- (d) On peut envisager de nombreux aspects. Aussi, pénalisez toute idée inadéquate.

Décomptez :

- 1 point pour une mauvaise rédaction de l'adresse.
- 1 point pour un manque de cohérence entre la vedette et les salutations.
- 1 point si l'élève a signé « M. John Smith » ou « Mademoiselle Jeannette Smith ».
- 1/2 point pour d'autres erreurs de disposition ou de ponctuation, ou d'orthographe dans des mots essentiels.

Tous ces sujets fournissent un matériau suffisant pour une demi-heure de travail. On attend un minimum de 200 mots. Le choix est riche.

EPREUVE II — Maximum 50 points.

Question 1 (maximum 16 points).

Les points attribués à cette question sont habituellement beaucoup plus bas que pour les autres. Les examinateurs sont priés de ne pas considérer 11 comme le maximum attribuable au résumé.

A. — Attribuer un maximum de 2 points pour chacun des aspects suivants. Pour obtenir 2 points, le candidat doit avoir clairement compris l'idée et l'avoir exprimée correctement. Nuancez vos notes par 1 1/2, 1 ou 1/2 point. A la fin du travail, comptez une nouvelle fois l'ensemble en fonction de la fluidité et de la cohérence du résumé complet. Si un passage est incohérent, supprimez au maximum 1/4 des points attribués ; s'il est plutôt rocailleux, soustrayez 1/8. Cette correction devrait apparaître sur la composition en écrivant par exemple 10 — 1 = 9. Toute soustraction de points destinée à pénaliser un texte plus long que la limite fixée ou une proposition non seulement inexacte mais absurde doit apparaître séparément. Le total final doit être entouré d'un cercle dans la marge.

- 1° L'homme de la rue accède aux découvertes scientifiques,
- 2° sans douter de leur origine et de leurs effets.
- 3° et 4°. La demande de nouvelles découvertes au niveau de vie ne se ralentit pas.
- 5° Les hommes ont confiance en la science et son travail,
- 6° et croient qu'on ne peut arriver à rien sans elle.
- 7° Bien qu'ils reconnaissent que les découvertes ne sont pas toujours d'accord avec les idées reçues, les nouvelles découvertes, les hommes restent convaincus que le public est convaincu que la science est la seule voie pour trouver l'unanimité ou la vérité.

B. — La limite de 120 mots est absolue. Supprimez un point pour chaque mot en excès du maximum. Ne comptez pas les mots productifs tels que « Dans ce passage... ». Recomptez les mots. N'accusez pas le candidat inscrit par le candidat.

Question 2 (maximum 20 points).

En notant cette question, le candidat choisit la section au candidat qui est le plus intéressante.

- (a) (1) Ceux qui aiment que l'on fasse des découvertes (2) et ceux qui préfèrent que l'on fasse un plan général permettant de résoudre le problème au moment (esprit d'exploration).
- (b) Visiter des grottes. A moins que vous ne soyez un géologue, vous ne connaissez pas très bien le terrain. Respectez toutes les règles.
- (c) (1) Il ne faut guère d'équipement pour faire des découvertes à petite échelle.
- (d) Quand le géologue casse une pierre, c'est un acte unique, car personne ne peut recommencer à casser à nouveau cette même pierre.
- (e) (1) Mesures de sécurité imposées ou recommandées. (2) Pour comprendre exactement les conséquences de ces mesures.

s).

re et logique.

s aspects de la question. On
ion rapportée ; on n'exigera
t particulier.

tion de l'objet et une explica-
e punissez pas sévèrement les
plement la clarté de l'expres-

ux aspects. Aussi, pénalisez

se rédaction de l'adresse.
le cohérence entre la vedette

« M. John Smith » ou « Made-
nith ».

erreurs de disposition ou de
orthographe dans des mots

n matériau suffisant pour une
un minimum de 200 mots. Le

s).

question sont habituellement
utres. Les examinateurs sont
omme le maximum attribuable

de 2 points pour chacun des
points, le candidat doit avoir
oir exprimée correctement.
1/2 point. A la fin du travail,
ble en fonction de la fluidité
mplet. Si un passage est in-
1/4 des points attribués ; s'il
1/8. Cette correction devrait
crivant par exemple 10 — 1
destinée à pénaliser un texte
e proposition non seulement
tre séparément. Le total final
a marge.

- 1° L'homme de la rue accepte aujourd'hui les découvertes scientifiques,
- 2° sans douter de leur origine, de leur validité, ou de leurs effets.
- 3° et 4°. La demande de nouveaux progrès destinés à élever le niveau de vie ne se ralentit jamais.
- 5° Les hommes ont confiance dans l'homme de science et dans son travail,
- 6° et croient qu'on ne peut arrêter le progrès.
- 7° Bien qu'ils reconnaissent que les hommes de science ne sont pas toujours d'accord sur les sécurités apportées par les nouvelles découvertes,
- 8° le public est convaincu que les hommes de science finiront par trouver l'unanimité ou au moins un large accord.

B. — La limite de 120 mots donne une marge généreuse. Supprimez un point pour chaque tranche de 5 mots dépassant le maximum. Ne comptez pas dans le comptage des mots introductifs tels que « Dans ce passage, l'auteur explique que... ». Recomptez les mots. N'acceptez pas simplement le chiffre inscrit par le candidat.

Question 2 (maximum 20 points).

En notant cette question, accordez le maximum dans chaque section au candidat qui expose nettement le point.

- (a) (1) Ceux qui aiment que l'on prévienne tout à leur place (1)
(2) et ceux qui préfèrent disposer simplement d'un plan général permettant de suivre l'inspiration du moment (esprit d'exploration) (2)
- (b) Visiter des grottes. (1)
A moins que vous ne soyez guidé par quelqu'un qui connaît très bien le terrain. (1)
Respectez toutes les règles de sécurité. (2)
- (c) (1) Il ne faut guère d'équipement spécial. (2)
(2) Le géologue fait continuellement de nouvelles découvertes à petite échelle. (1 1/2)
(1/2)
- (d) Quand le géologue casse une pierre, il pose, en fait, un acte unique, car personne après lui ne pourra casser à nouveau cette même pierre. (2)
- (e) (1) Mesures de sécurité imposées ou recommandées. (1)
(2) Pour comprendre exactement les conséquences de ce qu'on fait. (1)

- (f) Souterrain :
qui est sous le sol, sous la surface de la terre ; (1)
sous terre. (1)

Spéléologues :

personnes qui étudient les grottes scientifique-
ment ; explorateurs de grottes. (1)

Consciencieusement :

de bonne foi, sans se laisser distraire, avec
beaucoup d'application (1)
honnêtement (1/2)
fidèlement (1/2)

Etc.

Pour les sous-questions (a), (b) et (e), n'attribuez aucun point aux élèves qui se bornent à recopier une partie du texte.

Question 3 (maximum 6 points).

- (a) Attribuez un point pour toute phrase construite correctement.

N'attribuez aucun point en cas de construction incorrecte.

- (b) Accordez un point pour chaque définition exacte exprimée dans une phrase correcte (maximum 3 points).

Accordez un point pour chaque phrase où le sens du mot commençant par le même préfixe ressort clairement (maximum 3 points).

1/2 point pour une définition exprimée en une phrase incorrecte.

1/2 point pour une définition fournie en phrase incomplète.

Aucun point si le mot commençant par le même préfixe n'est pas présenté dans une phrase.

Un *monopole* est une propriété exclusive détenue par une firme (ou)

Un *monopole* est le nom donné à une firme qui détient des droits commerciaux exclusifs.

Un *interlude* est un intervalle ménagé au cours de la représentation d'une pièce (ou un intervalle dans le déroulement d'un événement).

Etc.

Question 2 (maximum 8 points)

- (a) Accordez un point par phrase.

Vérifiez la présence des id

N'accordez aucun point si aucune phrase ne contient
quelle phrase le candidat a

- (b) Les candidats répondront

Déduisez 1 point par faute de
toute construction boiteuse
incorrecte, 1/2 point pour
d'une virgule *essentielle*,

Revoyez la note totale en fin
du passage.

Pour les questions 1 et 2
des points, arrondissez à l'ent

Si les deux questions de
point, arrondissez l'un par

Mêmes remarques pour le

la surface de la terre ; (1)
(1)
s grottes scientifique-
ottes. (1)
laisser distraire, avec
(1)
(1/2)
(1/2)
b) et (e), n'attribuez aucun
recopier une partie du texte.
phrase construite correcte-
s de construction incorrecte.
e définition exacte exprimée
mum 3 points).
e phrase où le sens du mot
ixe ressort clairement (maxi-
primée en une phrase incor-
urnie en phrase incomplète.
nçant par le même préfixe
ase.
é exclusive détenue par une
à une firme qui détient des
énagé au cours de la repré-
tervalle dans le déroulement

Question 2 (maximum 8 points).

(a) Accordez un point par phrase.

Vérifiez la présence des idées suivantes...

N'accordez aucun point si l'on ne voit pas clairement à quelle phrase le candidat se réfère.

(b) Les candidats répondront de façons différentes.

Déduisez 1 point par faute d'orthographe, 2 points pour toute construction boiteuse, 1 point pour une expression incorrecte, 1/2 point pour l'omission ou l'emploi erroné d'une virgule *essentielle*, 1 point pour chaque idée omise.

Revoyez la note totale en fonction de l'impression générale du passage.

Pour les questions 1 et 2, si le candidat obtient la moitié des points, arrondissez à l'unité supérieure (ex. : $6 \frac{1}{2} = 7$).

Si les deux questions donnent une note comprenant 1/2 point, arrondissez l'un par excès et l'autre par défaut.

Mêmes remarques pour les questions 3 et 4.

EXEMPLE D'ENSEIGNEMENT SEMI-INDIVIDUALISE. ¹

La New Trier Township High School, Winnetka.

La *New Trier Township High School* est une grande école du degré secondaire supérieur où l'enseignement est semi-individualisé et où l'étudiant peut corriger son orientation jusqu'au terme de l'adolescence.

Elle accueille les élèves à partir de 14 ans ² et jouit d'une grande réputation tant aux Etats-Unis qu'à l'étranger.

3 740 étudiants suivent régulièrement les cours ³. Le corps professoral compte 255 membres dont :

40	professeurs d'anglais
31	» de mathématiques
30	» d'éducation physique
29	» de langues étrangères
27	» de « Social Studies » (Histoire, géographie, civisme, sociologie et économie)
17	» de sciences
9	» de musique
7	» de cours techniques
7	» de commerce
6	» d'art dramatique
6	» de peinture-dessin
6	» d'automobile
4	» d'économie domestique
2	» d'hygiène
1	» d'enseignement spécial (retardés mentaux édu-cables).

On compte un enseignant pour 15-16 étudiants et les classes réunissent généralement 25 élèves environ.

¹ Extrait d'un article que nous avons publié dans *Education*, en mai 1961.

² On sait que l'organisation de l'enseignement des Etats-Unis n'est pas uniforme. La N.T. High School relève du système « NK 8-4 » : un an de Nursery school, un an de jardin d'enfants, 8 ans de primaire et 4 ans de secondaire. Les deux autres systèmes les plus fréquents sont : NK 6-3-3 et NK 6-6. On rencontre aussi NK 7-5, NK 6-2-4 et NK 6-4-4.

³ Ces chiffres se réfèrent à l'année scolaire 1959-1960. Ils sont extraits de : *Information for College Admission Officers*, N.T. Township High School, nov. 1959, ou ont été recueillis sur place.

Les prestations d'un profes-sabilités spéciales (telles que ex.) comportent 24 périodes de de 5 périodes et 4 périodes de

Le niveau intellectuel des tests classiques, est élevé. En leurs études à New Trier obtie la moyenne nationale, dans les sances (SCAT, STEP, *Nation* continuent des études supérieu

L'individualisation des program

Le principe fondamental d Trier High School est défini programme de cours : « Person hommes sans éprouver un pla variété infinie. » ¹

Restant donc ainsi imprég netka, les études sont organisé possibilité de se développer capacités et ses penchants. C fantaisie stérile ou que les sol

Selon le système répandu peut obtenir un diplôme de celles-ci, il a gagné un nomb fixé par le conseil d'administr compte de normes générales unité conventionnelle l'impor des différents cours, pour un s par exemple, à un semestre cinq périodes par semaine). L l'étudiant atteint une note s fixée.

Pour obtenir le certificat School, il faut y avoir gagné

Le jeu des *credits* est à sécurité, car il permet de de liberté de l'élève.

En principe, celui-ci établi on compose le menu d'un re ce qui attire et convient le mie

¹ NEW TRIER TOWNSHIP HIGH SCH

III.

SEMI-INDIVIDUALISE. ¹

ool, Winnetka.

School est une grande école
où l'enseignement est semi-
corriger son orientation jus-

tir de 14 ans ² et jouit d'une
Unis qu'à l'étranger.

rement les cours ³. Le corps
nt :

ie
es
» (Histoire, géographie, ci-
économie)

que
cial (retardés mentaux édu-

ur 15-16 étudiants et les
lèves environ.

ans Education, en mai 1961.
des Etats-Unis n'est pas uniforme. La
un an de Nursery school, un an de
secondaire. Les deux autres systèmes
rencontre aussi NK 7-5, NK 6-2-4 et

9-1960. Ils sont extraits de : Information
School, nov. 1959, ou ont été recueil-

Les prestations d'un professeur qui n'assure pas de respon-
sabilités spéciales (telles que président d'un département, par
ex.) comportent 24 périodes de 40 minutes par semaine : 4 cours
de 5 périodes et 4 périodes de *counseling*.

Le niveau intellectuel des étudiants, jugés sur la base des
tests classiques, est élevé. Environ 80 % de ceux qui terminent
leurs études à New Trier obtiennent des résultats supérieurs à
la moyenne nationale, dans les tests d'aptitudes et de connais-
sances (SCAT, STEP, *National Merit*). 92 % des diplômés
continuent des études supérieures.

L'individualisation des programmes.

Le principe fondamental de l'action pédagogique de la New
Trier High School est défini dans la première phrase de son
programme de cours : « Personne ne peut croire à la dignité des
hommes sans éprouver un plaisir profond au spectacle de leur
variété infinie. » ¹

Restant donc ainsi imprégnées de l'esprit du Plan de Win-
netka, les études sont organisées de façon à fournir à chacun la
possibilité de se développer à son rythme propre, selon ses
capacités et ses penchants. Ceci ne signifie nullement que la
fantaisie stérile ou que les solutions de facilité soient permises.

Selon le système répandu dans tout le pays, l'étudiant ne
peut obtenir un diplôme de fin d'études que si, au cours de
celles-ci, il a gagné un nombre total de points ou de *credits*
fixé par le conseil d'administration de l'école qui tient lui-même
compte de normes générales. Ces *credits* expriment en une
unité conventionnelle l'importance qualitative et quantitative
des différents cours, pour un semestre (5 *credits* correspondent,
par exemple, à un semestre de cours « mineur », à raison de
cinq périodes par semaine). Les *credits* ne sont acquis que si
l'étudiant atteint une note supérieure à une limite minimum
fixée.

Pour obtenir le certificat d'études de la New Trier High
School, il faut y avoir gagné 350 *credits* en 4 ans.

Le jeu des *credits* est à la fois source de souplesse et de
sécurité, car il permet de délimiter exactement le champ de
liberté de l'élève.

En principe, celui-ci établit son programme d'études comme
on compose le menu d'un repas en choisissant, sur la carte,
ce qui attire et convient le mieux.

¹ NEW TRIER TOWNSHIP HIGH SCHOOL, *Curriculum Guide*, déc. 1959, p. 1.

Voici, par exemple, la liste des cours offerts pour la première année, avec mention des *credits* qu'ils rapportent s'ils sont suivis avec succès pendant un semestre¹ :

Cours « majeurs »

(Rapportant 10 *credits* par semestre ; sauf indication contraire, ils doivent être suivis pendant 2 semestres consécutifs)

Anglais	
Algèbre	Mathématiques générales
Alimentation (1 semestre)	Peinture-dessin (2 périodes par jour)
Civisme	Photographie
Commerce	Radio amateur
Dessin industriel	Travail du bois
Electricité (1 semestre)	Travaux manuels
Histoire universelle	Vêtement (1 semestre)
Langues étrangères	

Cours « mineurs »

(Nombre de *credits* entre parenthèses)

Art dramatique (6)	Dessin industriel (1 période par jour) (5)
Chant choral (3)	Diction (6)
Dactylographie (5)	Harmonie (5)
Cours d'harmonie (6)	Orchestre symphonique (5)
	Travaux manuels (1 période par jour) (5)

Néanmoins, plusieurs restrictions influencent le choix de l'élève. En premier lieu, un certain nombre de cours, jugés indispensables à la culture de base de tous les membres de la nation, sont obligatoires :

<i>Cours</i>	<i>Durée obligatoire</i>	<i>Credits</i>
Langue maternelle	4 ans	80
Mathématiques	2 ans	40
« Social Studies »	2 ans ²	40
Sciences	1 an	20
Education physique	4 ans	16
Automobile : théorie et pilotage	1 semestre	3
		Total 199

¹ Cf. NEW TRIER TOWNSHIP HIGH SCHOOL, *Registration Bulletin for Freshmen 1959-1960. Planning a Course of Study*, p. 2.

² Dont un an obligatoirement consacré à l'histoire des Etats-Unis.

Environ 2/3 du total des *credits* des cours imposés. Toutefois, être satisfait à ces exigences nous avons relevé une dizaine de *credits* d'accomplir valablement l'année.

Un second facteur important de son programme : la *réputation* aux quelles il aspire. En *premier lieu* mine ses conditions d'admission bien de *credits* l'étudiant doit secondaire, pour des branches.

Enfin, les parents, les *maîtres* et les orienteurs s'efforcent de *répondre* à ses intérêts, veillant à ce qu'il tire *le meilleur* de ses potentialités.

La semaine scolaire compte *vingt-cinq* minutes, — réunion quotidienne (20') et durée du lunch toujours *deux* heures, — on suggère généralement *un* programme moyennement chargé *pour* l'étude personnelle. Voici *quelques* exemples de la première année¹ :

I		II	
<i>Cours</i>	<i>pér. sem.</i>	<i>Cours</i>	<i>pér. sem.</i>
Anglais	5	Anglais	5
Algèbre	5	Algèbre	5
Latin	5	Sciences	5
Peint./dess.	10	Commerce	5
Gymn. (filles)	4	Gymn. (garç.)	5
	29	Musique instr.	5
Etude personnelle	11		30
	40		10
			40

On remarquera combien *facile* avec l'émission de l'effort *journalier* ment, tous les cours académiques *peuvent* être faits la matinée.

¹ NEW TRIER TOWNSHIP HIGH SCHOOL

des cours offerts pour la première année qu'ils rapportent s'ils sont suivis pendant un semestre¹ :

« Cours »

sauf indication contraire, les cours sont exigés pendant 2 semestres consécutifs)

Mathématiques générales
 Dessin (2 périodes par jour)
 Photographie
 Musique amateur
 Travail du bois
 Travaux manuels
 Travail manuel (1 semestre)

« Cours »

(entre parenthèses)

Industriel (1 période par jour) (5)
 (6)
 Symphonique (5)
 Travaux manuels (1 période par jour) (5)

Les facteurs suivants influencent le choix de la durée et du nombre de cours, jugés acceptables par tous les membres de la commission :

Durée obligatoire	Credits
4 ans	80
2 ans	40
2 ans ²	40
1 an	20
4 ans	16
1 semestre	3
Total	199

1. Registration Bulletin for Freshmen
 des Etats-Unis.

Environ 2/3 du total des *credits* exigés sont donc fournis par des cours imposés. Toutefois, comme nous le verrons, il peut être satisfait à ces exigences de façon fort libre ; ainsi, nous avons relevé une dizaine de possibilités différentes permettant d'accomplir valablement l'année de sciences réclamée.

Un second facteur important guide l'étudiant dans l'élaboration de son programme : la profession ou les études supérieures auxquelles il aspire. En particulier, chaque université détermine ses conditions d'admission et spécifie notamment combien de *credits* l'étudiant doit avoir acquis dans l'enseignement secondaire, pour des branches déterminées.

Enfin, les parents, les maîtres, les conseillers pédagogiques et les orienteurs s'efforcent de guider l'étudiant au mieux de ses intérêts, veillant à ce qu'il tire le meilleur profit possible de ses potentialités.

La semaine scolaire comptant 5 jours de 8 périodes de 40 minutes, — réunion quotidienne avec le conseiller pédagogique (20') et durée du lunch toujours pris à l'école (25') non comprises, — on suggère généralement à l'élève de se constituer un programme moyennement chargé, ménageant du temps pour l'étude personnelle. Voici quatre exemples types pour la première année¹ :

I		II		III		IV	
Cours	pér. sem.	Cours	pér. sem.	Cours	pér. sem.	Cours	pér. sem.
Anglais	5	Anglais	5	Anglais	5	Anglais	5
Algèbre	5	Algèbre	5	Algèbre	5	Algèbre	5
Latin	5	Sciences	5	Histoire	5	Histoire	5
Peint./dess.	10	Commerce	5	Aliment.	7	Sciences	5
Gymn. (filles)	4	Gymn. (garç.)	5	Chant	3	Lang. étrang.	5
	—	Musique instr.	5	Gymn. (filles)	4	Gymn. (garç.)	5
	29		—		—		—
Etude personnelle	11		30		29		30
	—		10		11		10
	40		—		—		—
			40		40		40

On remarquera combien ces plans de travail contrastent avec l'émiettement de l'effort si fréquent chez nous. Pratiquement, tous les cours académiques peuvent être concentrés dans la matinée.

¹ NEW TRIER TOWNSHIP HIGH SCHOOL, *Courses for Freshmen*, avril 1959, p. 4.

Environ un quart du temps passé à l'école est réservé aux études personnelles. Les heures ainsi laissées libres seront souvent consacrées à des travaux de recherche dans la bibliothèque scolaire qui, aux Etats-Unis, joue un rôle incomparablement plus important qu'en Belgique.

L'individualisation de l'enseignement.

Non seulement l'élève choisit les branches qui lui conviennent le mieux, mais encore l'enseignement de chacune de celles-ci sera adapté à ses possibilités.

En effet, toutes les branches importantes peuvent être étudiées à cinq niveaux d'aptitudes différents : inférieur, moyen faible, normal, accéléré, avancé. De cette façon, l'effort réclamé à l'étudiant, qu'il possède une intelligence supérieure ou soit peu doué, qu'il soit fort dans un domaine et en retard dans un autre, est toujours en proportion avec ses possibilités.

Comme l'élève suit rarement tous ses cours à un même niveau, le danger d'une ségrégation générale, selon les aptitudes, semble minime : l'école reste d'ailleurs très attentive à ce problème, regroupant systématiquement tous les étudiants à l'occasion de certaines activités. On met aussi tout en œuvre pour inculquer une véritable tolérance vis-à-vis du plus ou moins grand talent des compagnons d'études et pour encourager chacun à se dépasser : « En sport, tout le monde n'est pas capable de jouer en excellence ; c'est aussi le cas dans les études. Mais on attend de chacun le meilleur de lui-même.. »¹

Le système d'enseignement à différents niveaux est pratiqué à New Trier, depuis la fin de la première guerre mondiale, à la plus grande satisfaction de tous. En moyenne, les étudiants se répartissent selon les pourcentages suivants :

— supérieurs	15—20 %
— normaux (moyens forts)	40—55 %
— moyens faibles	36—40 %
— limités	5— 8 %
— avancés (seniors)	± 10 %

Quand les élèves entrent à la New Trier High School, les groupes sont provisoirement déterminés sur la base des résultats scolaires antérieurs et d'autres renseignements réunis selon un système que nous étudierons plus loin. Par après, les résultats obtenus dans l'école même corrigeront et guideront les affectations, celles-ci n'étant jamais définitives, quelle que soit la branche.

1 NEW TRIER HIGH SCHOOL, *Guide Book to New Trier*, Winnetka, 1959, p. 39.

Pareille souplesse conduira à un système plus complexe, mais il ne faut cependant pas se précipiter.

Dans le tableau suivant, nous indiquons les possibilités offertes pendant l'année scolaire 1959-1960. Quelques-unes de ces possibilités sont des études¹.

- Chaque cours est désigné par un chiffre :
- a) Le chiffre des centaines indique le nombre de semestres qu'on aura donc 1, 2, 3 et 4.
 - b) Le chiffre des dizaines indique le niveau de l'étudiant :
 - 1 = 1^{er} semestre
 - 2 = 2^e semestre
 - 0 = peut être
 - 3 = cours de
 - c) Le chiffre des unités indique le niveau de l'étudiant :
 - 1 = niveau inférieur
 - 2 = moyen faible
 - 3 = normal
 - 4 = enseigné
 - 5 = avancé
 - 6 = séminaire
 - 9 = tous niveaux

Branches et niveaux de cours

Langue maternelle
125

124
123

122

122 R

121

224 - 223 - 222 - 221
324 - 323 - 322 - 321
424 - 424 (+) - 423 - 422 - 421

1 Les deux documents de base sur lesquels nous nous appuyons sont : NEW TRIER HIGH SCHOOL, *Curriculum Guide*, 1959.

sé à l'école est réservé aux
ainsi laissées libres seront
le recherche dans la biblio-
joue un rôle incomparable-

nt.

s branches qui lui convien-
nement de chacune de cel-

importantes peuvent être
différents : inférieur, moyen
cette façon, l'effort réclamé
intelligence supérieure ou soit
maine et en retard dans un
c ses possibilités.

ous ses cours à un même
générale, selon les aptitu-
'ailleurs très attentive à ce
ment tous les étudiants à
n met aussi tout en œuvre
ance vis-à-vis du plus ou
d'études et pour encoura-
rt, tout le monde n'est pas
est aussi le cas dans les
e meilleur de lui-même.»¹

différents niveaux est prati-
première guerre mondiale,
us. En moyenne, les étu-
rcentages suivants :

15—20 %
arts) 40—55 %
36—40 %
5—8 %
± 10 %

ew Trier High School, les
nés sur la base des résul-
enseignements réunis selon
loin. Par après, les résul-
rrigeront et guideront les
définitives, quelle que soit

New Trier, Winnetka, 1959, p. 39.

Pareille souplesse conduit certes à une organisation com-
plexe, mais il ne faut cependant pas exagérer la difficulté.

Dans le tableau suivant, nous faisons apparaître toutes les
possibilités offertes pendant le second semestre de l'année
scolaire 1959-1960. Quelques notes marginales précisent l'esprit
des études¹.

Chaque cours est désigné par un nombre de trois chiffres :

- a) Le chiffre des centaines indique en quelle année il se donne ;
on aura donc 1, 2, 3 et 4.
- b) Le chiffre des dizaines indique le semestre :
- 1 = 1^{er} semestre
 - 2 = 2^e semestre
 - 0 = peut être suivi au 1^{er} ou au 2^e semestre
 - 3 = cours de vacances d'été.
- c) Le chiffre des unités indique le niveau d'aptitude :
- 1 = niveau inférieur
 - 2 = moyen faible
 - 3 = normal
 - 4 = enseignement accéléré
 - 5 = avancé
 - 6 = séminaires
 - 9 = tous niveaux réunis.

Branches et niveaux de cours	Observations
Langue maternelle	
125	— Pour étudiants supérieurement intel- ligents, excellents en langue mater- nelle (intègre la langue maternelle, la biologie et l'histoire)
124	— Cours enrichi, fondé sur les œuvres littéraires de niveau universel
123	— Cours moyen
122	— Cours moyen inférieur ; s'adresse plus particulièrement aux élèves qui n'ont pas encore acquis une métho- de de travail rationnelle
122 R	— R = « remedial ». Destiné aux élè- ves identifiés par le bureau de testing comme présentant une défi- cience marquée en lecture (com- préhension et rapidité) et en ortho- graphe.
121	— Accueille les élèves présentant une faiblesse générale dans toutes les branches de la langue maternelle.
224 - 223 - 222 - 221	Les différences de niveaux en 2 ^e , 3 ^e et 4 ^e années sont <i>mutatis mutandis</i> parallèles à celles que nous indiquons ci-dessus pour la 1 ^{re} . 323 = journalisme. 424 (+) = « Great Books ».
324 - 323 - 322 - 321	
424 - 424 (+) - 423 - 422 - 421	

¹ Les deux documents de base suivants ont été utilisés : NEW TRIER TOWNSHIP
HIGH SCHOOL, *Curriculum Guide*, 1959 et *Program of Classes*, 1959-1960.

Branches et niveaux de cours	Observations
Mathématiques	<p>La différence entre les niveaux 2, 3 et 4 réside moins dans l'accélération que dans la méthode d'enseignement et l'approfondissement.</p> <p><i>Au niveau 2</i>, les explications sont détaillées; beaucoup d'applications; pas d'incursions dans les domaines voisins.</p> <p><i>Au niveau 3</i>, l'étudiant doit plus travailler par lui-même: théorie plus rigoureuse, enrichissement.</p> <p><i>Au niveau 4</i>, les concepts sont traités rapidement; étude plus approfondie; nombreuses incursions dans les domaines voisins.</p> <p>Le même manuel est employé pour ces 3 niveaux et tous les élèves étudient le même chapitre en même temps, ce qui permet, à tout moment, le passage d'un niveau à l'autre.</p> <p>4 semestres de mathématiques sont obligatoires; toutefois, 2 de ceux-ci peuvent être consacrés à la comptabilité ou à l'arithmétique commerciale.</p>
Algèbre A (accélééré) 124	
» — 124	
» E (expérimental) 124	
» — 123	
» E 123	
» — 122	
» E 122	
Mathématiques de base 121	
Mathématiques approfondies (20 credits) 225	
Géométrie 224	
» 223	
» 222	
Mathématiques 222	
Mathématiques de base 221	
Mathématiques approfondies (20 credits) 315	
Mathématiques 324	
Algèbre 323	
» 322	
Géométrie 303	
Mathématiques approfondies 425	
Mathématiques 424	
Algèbre (niveau universitaire) 403	
Trigonométrie 403	
Usage de la règle à calculs 323 (2 credits; va avec algèbre 322)	
« Social studies »	<p>4 semestres de « Social Studies » sont obligatoires; 2 de ceux-ci doivent être consacrés à l'histoire des Etats-Unis.</p>
Civisme 123, 122, 121	
Histoire universelle 125, 124, 123	
Histoire: temps modernes 224	
Histoire: antiquité 204	
Histoire: moyen âge 202, 203, 204	
Histoire: universelle 221	
Géographie 222 - 223	
Histoire: temps modernes 323	
Histoire: Grande-Bretagne 303	
Histoire: U.S.A. 324 - 323 - 322 - 321	
Histoire: U.S.A. (approfondie) 425	
Histoire: contemporaine (Europe) 425	
Histoire: Grande-Bretagne (approf.) 404	
Histoire: Amérique latine 403	
Histoire: Extrême-Orient 403	
Histoire: Universelle, XX ^e S. 402	
Histoire: Etats-Unis, XX ^e S. 402	
Civisme 403	
Sociologie 403	
Sciences économiques 403	

Branches et niveaux de cours	
Sciences	
Biologie	123 - 124 - 125
Radio amateur	123
Biologie	224 - 223 - 222
Chimie	225
Biologie	323 - 322 - 321
Electronique	323
Physique	325
Chimie	324 - 323
Chimie (séminaire)	426
Physique	425 - 424 - 423
Sciences	409
Langues	
Latin	
	124 - 123 - 122
	224 - 223 - 222
	324 - 323
	424 - 423
Allemand	124 - 123
	224 223 223C
	324 - 323
	426
Russe	124
	224
	324
Français	125 124 123
	224C 224 223C 223
	324C 324 323C 323 324 =
	424C 423C
Espagnol	125 124 123 122
	224 224C 223 223C 222
	324 324 = 323 322
	424 423
Arts	
Dessin-peinture	129
Travail manuel	121
Dessin-peinture	229C 229
Travail manuel	221
Dessin-peinture	329
Céramique	329
Joaillerie	329
Histoire de l'art	329
Céramique	429
Joaillerie	429
Peinture-dessin	429

Observations

différence entre les niveaux 2, 3 réside moins dans l'accélération dans la méthode d'enseignement qu'au approfondissement.

niveau 2, les explications sont plus développées; beaucoup d'applications; de nombreuses incursions dans les domaines connexes.

niveau 3, l'étudiant doit plus travailler par lui-même: théorie plus rigoureuse, enrichissement.

niveau 4, les concepts sont traités plus complètement; étude plus approfondie; nombreuses incursions dans les domaines connexes.

le même manuel est employé pour les deux niveaux et tous les élèves étudient le même chapitre en même temps, ce qui permet, à tout moment, le passage d'un niveau à l'autre.

deux semestres de mathématiques sont obligatoires; toutefois, 2 de ceux-ci peuvent être consacrés à la comptabilité ou à l'arithmétique commerciale.

deux semestres de « Social Studies » sont obligatoires; 2 de ceux-ci doivent être consacrés à l'histoire des Etats-

Branches et niveaux de cours		Observations
Sciences		Les élèves qui désirent continuer des études dans les Facultés de Sciences des Universités sont invités à suivre 6 semestres de sciences et 8 de mathématiques.
Biologie	123 - 124 - 125	
Radio amateur	123	
Biologie	224 - 223 - 222	
Chimie	225	
Biologie	323 - 322 - 321	
Electronique	323	
Physique	325	
Chimie	324 - 323	
Chimie (séminaire)	426	
Physique	425 - 424 - 423	Aucun cours de langue n'est obligatoire. Les étudiants sont cependant fort encouragés à les suivre s'ils se destinent aux études supérieures. On pense que, très prochainement, les universités exigeront à l'entrée que l'étudiant ait étudié au moins une langue étrangère pendant 3 ans.
Sciences	409	
Langues		
Latin	124 - 123 - 122	
	224 - 223 - 222	
	324 - 323	
	424 - 423	
Allemand	124 - 123	
	224 223 223C	
	324 - 323	
	426	
Russe	124	C indique qu'il s'agit de la continuation d'un cours précédent. = cours de conversation.
	224	
	324	
Français	125 124 123	
	224C 224 223C 223	
	324C 324 323C 323 324=	
	424C 423C	
Espagnol	125 124 123 122	
	224 224C 223 223C 222	
	324 324= 323 322	
	424 423	
Arts		Tous ces cours peuvent être suivis à raison de 2 périodes par jour. (= cours « majeur » : 10 <i>credits</i>) ou de 1 période par jour (= cours « mineur » : 5 <i>credits</i>). Toutefois, Histoire de l'art 329 est toujours un « majeur ».
Dessin-peinture	129	
Travail manuel	121	
Dessin-peinture	229C 229	
Travail manuel	221	
Dessin-peinture	329	
Céramique	329	
Joaillerie	329	
Histoire de l'art	329	
Céramique	429	
Joaillerie	429	
Peinture-dessin	429	

Branches et niveaux de cours		Observations
Commerce		Objectifs poursuivis par ce cours : 1. Préparation aux études supérieures de sciences commerciales ; 2. Formation générale ; 3. Préparation pour les étudiants qui veulent travailler part-time dans le commerce pendant leurs études universitaires ; 4. Préparation professionnelle pour les élèves qui ne feront pas d'études supérieures ; 5. Possibilité pour les étudiants qui cherchent leur voie, de voir si la carrière commerciale pourrait les intéresser. Le 422 = une réunion de 30' tous les matins à l'école ; le reste consiste en stages pratiques (20 h/semaine).
Commerce général	122 123	
Commerce général	222	
Comptabilité	223	
Le point de vue du consommateur 323		
Vente	302	
Publicité	303	
Organisation commerciale	303	
Sténographie	323	
Dactylographie	303	
Pratique du bureau	322	
Secrétariat	423	
Sténographie	423	
Droit commercial	423 403	
Pratique commerciale	422	
Automobile		Une loi, passée en 1955, rend ce cours obligatoire dans toutes les High Schools de l'Illinois. (3 « credits »).
Théorie et pratique (pilotage)	209	
Economie domestique		Destiné principalement aux jeunes filles ; toutefois : — les garçons s'intéressent à la restauration ou à l'hôtellerie peuvent suivre les cours d'alimentation ; — les garçons s'intéressant à l'architecture peuvent suivre les cours de décoration intérieure.
Alimentation	109	
Vêtement	109	
Alimentation	209	
Vêtement	209	
Décoration intérieure	329	
Cours techniques		— Le but poursuivi n'est pas de donner une formation professionnelle, mais de faire de l'étudiant « un consommateur intelligent des produits de l'industrie » (p. 64). — Chacun de ces cours rapporte 5 ou 10 « credits » selon le temps qu'on leur consacre. — Les universités reconnaissent ces « credits » ; de plus en plus, elles souhaitent que les étudiants acquièrent une formation technique.
Dessin industriel	123 122	
Bois	123 121	
Electricité	103	
Dessin (architecture)	223	
Bois	222	
Métallurgie générale	223	
Dessin (mécanique)	323	
Métallurgie générale	323	
Notions d'architecture	323	
Lecture de plans	302	

Branches et niveaux de cours	
Musique	
Chant	1
Orchestre	1
Cours d'harmonie	1
Appréciation musicale	1
Chant	
Petit groupe vocal	2
Appréciation musicale	2
Composition	
Madrigaux (groupe vocal)	3
Composition	3
Chorale (sélection de 32 étud.)	
Opéra	0
Grande chorale	0
Harmonie (cadets - groupe de sorties - groupe d'honneur)	0
Orchestre (seniors)	0
Ensemble instrumental	0
Piano	0
Education physique - Sports -	
Art oratoire - Art dramatique	
Photographie	
Croix-Rouge	
L'étude de chaque élève	
<p>Nous avons dit que l'école s'intéresse sur ses futurs élèves et leur offre immédiatement une place, dans un environnement complexe que nous venons de décrire.</p> <p>La majorité des étudiants viennent de six écoles publiques des alentours.</p> <p>Dès le mois de janvier, les <i>Schools</i> fournissent aux six districts les renseignements, d'aptitudes et de caractéristiques des candidats qui ont manifesté l'intérêt pour le <i>New Trier</i>. Les élèves des <i>New Trier</i> et <i>Winnetka</i>, en avril, pour y s...</p>	

Observations

tifs poursuivis par ce cours :
préparation aux études supérieures
sciences commerciales ;
formation générale ;
préparation pour les étudiants qui
préparent à travailler part-time dans le
commerce pendant leurs études uni-
versitaires ;
préparation professionnelle pour les
étudiants qui ne feront pas d'études
supérieures ;
opportunité pour les étudiants qui
cherchent leur voie, de voir si la
carrière commerciale pourrait les
intéresser.
= une réunion de 30' tous les
jours à l'école ; le reste consiste en
activités pratiques (20 h/semaine).

La loi, passée en 1955, rend ce
programme obligatoire dans toutes les High
Schools de l'Illinois. (3 « credits »).

est principalement aux jeunes
hommes ; toutefois :
- les garçons s'intéressant à la res-
tauration ou à l'hôtellerie peuvent
suivre les cours d'alimentation ;
- les garçons s'intéressant à l'archi-
tecture peuvent suivre les cours de
construction intérieure.

Un cours poursuivi n'est pas de don-
ner une formation professionnelle,
mais de faire de l'étudiant « un con-
sommateur intelligent des produits
de l'industrie » (p. 64).
Chaque un de ces cours rapporte 5 ou
6 « credits » selon le temps qu'on
y consacre.
Les universités reconnaissent ces
« credits » ; de plus en plus, elles
prévoient que les étudiants acqui-
eront une formation technique.

<i>Branches et niveaux de cours</i>	<i>Observations</i>
<i>Musique</i>	
Chant 129	Il s'agit soit d'une initiation pour l'amateur, soit d'études approfondies pouvant préparer aux carrières musi- cales.
Orchestre 129	
Cours d'harmonie 129	
Appréciation musicale 129	
Chant 229	
Petit groupe vocal 229	
Appréciation musicale 229	
Composition 329	
Madrigaux (groupe vocal) 429	
Composition 429	
Chorale (sélection de 32 étud.) 029	
Opéra 029	
Grande chorale 029	
Harmonie (cadets - groupe de sorties - groupe d'honneur) 029	
Orchestre (seniors) 029	
Ensemble instrumental 029	
Piano 029	

Education physique - Sports - Danse

Art oratoire - Art dramatique

Photographie

Croix-Rouge

L'étude de chaque élève avant son entrée à New Trier.

Nous avons dit que l'école recueille des informations détaillées sur ses futurs élèves afin que ceux-ci puissent trouver immédiatement une place, au moins provisoire, dans le système complexe que nous venons de décrire. Voici comment on procède.

La majorité des étudiants qui fréquentent New Trier proviennent de six écoles publiques et de sept écoles confessionnelles des alentours.

Dès le mois de janvier, les services de testing de la *High School* fournissent aux six écoles publiques des tests d'intelligence, d'aptitudes et de connaissances à administrer aux enfants qui ont manifesté l'intention de s'inscrire aux cours de New Trier. Les élèves des écoles confessionnelles viendront à Winnetka, en avril, pour y subir des épreuves similaires.

Chaque école doit en outre remplir une fiche individuelle qui porte, outre les résultats des tests mentionnés, un tableau résumant l'appréciation générale des professeurs¹ :

APPRECIATIONS GLOBALES

Intelligence	1	2	3	4	5
Application	1	2	3	4	5
Sens des responsabilités	1	2	3	4	5
Conduite à l'école	1	2	3	4	5
Qualités de chef	1	2	3	4	5

APPRECIATIONS SPECIALES

Anglais	1	2	3	4	5
Mathématiques :					
Raisonnement	1	2	3	4	5
Connaissances de base	1	2	3	4	5
« Social Studies »	1	2	3	4	5
Sciences	1	2	3	4	5

Code : 1 = supérieur ; 2 = au-dessus de la moyenne ; 3 = moyen ; 4 = en dessous de la moyenne ; 5 = pauvre. Le chiffre retenu est encadré.

Après l'administration des tests, le directeur du service de *counseling* de New Trier rencontre les conseillers des différentes écoles inférieures afin de recueillir des renseignements complémentaires concernant non seulement l'instruction, mais aussi l'histoire, la famille, la santé des futurs étudiants.

Ensuite, chacun de ceux-ci est interviewé dans son école. L'entretien porte d'abord sur les talents (arts, sports, bricolage) et sur les goûts (branches préférées), puis sur la famille, l'enquêteur tâchant d'identifier les problèmes éventuels. Une question choc termine l'entrevue.

Au mois de mai, parents et élèves — qui ont déjà pu étudier le programme des cours — sont invités à assister à une réunion d'information à la High School. Les inscriptions aux différents cours sont recueillies.

La direction de New Trier possède alors assez d'éléments pour constituer des groupes d'environ 30 étudiants qui, pendant toutes leurs études, seront confiés au même conseiller pédagogique et le rencontreront chaque matin. Dans ces groupes, l'hétérogénéité est systématiquement recherchée, tant en ce qui concerne les aptitudes que les origines socio-économiques. Toutefois, les sexes sont séparés, le conseiller étant toujours du même sexe que ses étudiants. De plus, on s'efforce de donner

¹ Cf. NEW TRIER HIGH SCHOOL, *Test and Personal Data Card*.

une personnalité équilibrée à ment que trop d'adolescents réunis.

Le jour de la rentrée, non res sont organisés, mais chacun qu'il doit faire : il sait, par exemple doit se joindre au groupe des groupe supérieur ; il sait qu'un déjà quels livres, quels cahiers acheter et on lui a fait connaître l'un de ces articles.

Terminons ces indications en signalant qu'environ 15 jours les parents sont invités à rencontrer le conseiller de leurs enfants les raisons du classement aux manifestations extra-

La semaine suivante, les parents de l'école avec leur fils et suivent à ses côtés (la journée est, à ce midi et la soirée). On devine que l'expérience facilite la compréhension.

Ayant commencé ses études dans ces conditions excellentes, l'étudiant, jour, par son conseiller.

Les conseillers pédagogiques.

Grâce à ses conseillers pédagogiques, le School réussit ce véritable tour de ment à guider efficacement chaque élève le dédale des programmes, mais un plan psychologique que sur le

Le *counseling* est assuré par 126 conseillers pédagogiques partant 1/5 de leur temps à cette mission, 10 conseillers full-time dirigés par 6 psychologues full-time.

Au cours de la réunion qu'un groupe de 30 étudiants dont nous rencontrons son conseiller pour discuter de son avenir, de carrière et, en général, de sa majorité. L'atmosphère détend et favorise une excellente transition vers la vie d'adulte.

emplir une fiche individuelle
ests mentionnés, un tableau
des professeurs¹ :

GLOBALES

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

SPECIALES

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

dessus de la moyenne ; 3 =
anne ; 5 = pauvre. Le chiffre

, le directeur du service de
les conseillers des différen-
ueillir des renseignements
eusement l'instruction, mais
les futurs étudiants.

interviewé dans son école.
ents (arts, sports, bricolage)
s), puis sur la famille, l'en-
èmes éventuels. Une ques-

s — qui ont déjà pu étudier
tés à assister à une réunion
inscriptions aux différents

ède alors assez d'éléments
n 30 étudiants qui, pendant
u même conseiller pédago-
matin. Dans ces groupes,
recherchée, tant en ce qui
igines socio-économiques.
conseiller étant toujours du
us, on s'efforce de donner

onal Data Card.

une personnalité équilibrée à chaque groupe en évitant notam-
ment que trop d'adolescents « à problèmes » ne s'y trouvent
réunis.

Le jour de la rentrée, non seulement les cours et les horai-
res sont organisés, mais chaque étudiant sait exactement ce
qu'il doit faire : il sait, par exemple, que pour telle branche, il
doit se joindre au groupe des moyens et pour telle autre, au
groupe supérieur ; il sait qu'un conseiller l'attend ; il sait même
déjà quels livres, quels cahiers, quel équipement sportif il doit
acheter et on lui a fait connaître d'avance le prix exact de cha-
cun de ces articles.

Terminons ces indications concernant les débuts scolaires
en signalant qu'environ 15 jours après la rentrée des classes,
les parents sont invités à une première réunion où ils rencon-
treront le conseiller de leurs enfants et où ils seront éclairés sur
les raisons du classement aux différents niveaux, sur la signifi-
cation des manifestations extra-scolaires, etc.

La semaine suivante, les pères et les mères se rendent à
l'école avec leur fils et suivent une journée de classe complète
à ses côtés (la journée est, à cette occasion, répartie sur l'après-
midi et la soirée). On devine aisément combien une telle expé-
rience facilite la compréhension entre l'école et la famille.

Ayant commencé ses études secondaires supérieures dans
ces conditions excellentes, l'étudiant sera suivi, jour après
jour, par son conseiller.

Les conseillers pédagogiques.

Grâce à ses conseillers pédagogiques, la New Trier High
School réussit ce véritable tour de force qui consiste, non seule-
ment à guider efficacement chacun de ses 3 740 étudiants dans
le dédale des programmes, mais aussi à les aider tant sur le
plan psychologique que sur le plan social et médical.

Le *counseling* est assuré par :

126 conseillers pédagogiques part-time, professeurs consacrant
1/5 de leur temps à cette mission.

10 conseillers full-time dirigeant les précédents.

6 psychologues full-time.

Au cours de la réunion quotidienne du matin, chaque grou-
pe de 30 étudiants dont nous avons signalé la constitution
rencontre son conseiller pour discuter de problèmes de discipli-
ne, de carrière et, en général, de tout ce qui peut intéresser la
majorité. L'atmosphère détendue de ces rencontres constitue
une excellente transition vers le travail ardu qui va bientôt
commencer.

L'étudiant peut aussi consulter individuellement son conseiller, chaque fois qu'il le désire.

De plus, pendant l'année, le conseiller rendra au moins une visite personnelle à la famille de chaque élève appartenant à son groupe.

Il va sans dire qu'une mission aussi délicate ne s'improvise pas : elle réclame des hommes ouverts aux problèmes de la jeunesse et spécialement formés pour l'aider.

Lors du recrutement des professeurs, le conseil d'administration de l'école attache une importance toute particulière aux qualités intellectuelles et morales indispensables au parrainage des étudiants et donne la priorité à ceux qui les possèdent.

Pendant sa première année de fonction, le conseiller doit assister à 50 réunions de formation d'environ 1/2 heure chacune et, après chaque cycle de 4 ans, il est soumis à un nouvel entraînement.

La préparation de base se déroule de la façon suivante :

I. Avant la rentrée des classes :

Un certain nombre de séances sont d'abord consacrées à l'étude de l'administration de l'école. Ensuite, le conseiller reçoit en communication toutes les informations recueillies sur ses futurs élèves.

Enfin, quelques jours avant la fin des vacances, le directeur des études de première année réunit les « freshman helpers », c'est-à-dire des étudiants de dernière année qui vont aider le conseiller dans sa tâche. Pendant les 9 premières semaines, l'assistant participera à toutes les réunions matinales. Pendant les 9 semaines suivantes, il ne fournira plus que deux prestations par semaine ; par après, il n'interviendra plus que sur demande du conseiller.

II. Pendant l'année scolaire.

Sous la présidence d'un directeur spécialisé, la formation se poursuit lors de causeries familières faites par les directeurs des départements (anglais, mathématiques, sciences, etc.), par les chefs des différents services (directeurs du testing, psychologue, bibliothécaire, médecin-assistant social, délégué à l'association des parents) et par les adultes responsables des clubs estudiantins.

Le directeur des conseillers convoque aussi certaines réunions spéciales pour envisager des problèmes nouveaux concernant l'admission d'élèves, la révision des programmes, etc.

Enfin, en avril, une conférence de clôture permet de dresser le bilan d'activité de l'année qui touche à sa fin et d'établir les premières prévisions pour l'année suivante.

AGAZZI, A., *Les aspects pédagogiques de l'Europe*, C.C.C., 1967.

BACHER, F., La normalisation de l'évaluation, numéro spécial de la revue *Revue de l'Éducation*, 1969, 131-156.

BACHER, F., L'évaluation des résultats, in *Le Travail humain*, 1968.

BAZIN, R., Les Français d'aujourd'hui, *Gestion*, 4, 1970, 3-10.

BLOCK, J., *Mastery Learning*, New York, 1968.

BLOOM, B., HASTING and MADSEN, *Student Learning*, New York, 1968.

BONNARDEL, R., Application de la méthode à l'étude de la notation des résultats, *Revue de l'Éducation*, 1946, 130-139.

BRITTON, J., Experimental Marking, Year-Olds, in *Educational Research*, 1968.

BRITTON, J., MARTIN, N. et ROSS, *An Account of an Experiment*, 1968.

BRUCE, G., *Secondary School Examinations*, Pergamon Press, 1969.

BRUNELLE, L., *Pourquoi des examens*, 1968.

CHOPPIN, B. et PURVES, A., Choice items dealing with the examination, *Teaching of English*, 3, 1, 1968.

DEMANGEON, M. et LARCEBEAUX, *Revue de l'Éducation*, 1958, 14, 131-156.

Docimologie et Évaluation, numéro spécial, 2-3, 1969, 166 p.

EDGEWORTH, F.V., *The Statistical Theory of the Examination*, *Stat. Society*, 1888, 51, 599-603.

Examen des examens, N° spécial, 1970.

Examinations Bulletins, Londres, 1968.

N° 1. *The Certificate of Secondary Education*, 1968.

N° 2. *The C.S.E. : Experiments*, 1968.

N° 3. *The C.S.E. : An Introduction*, 1968.

N° 4. *The C.S.E. : An Introduction*, 1964.

BIBLIOGRAPHIE

- AGAZZI, A., *Les aspects pédagogiques des examens*, Strasbourg, Conseil de l'Europe, C.C.C., 1967.
- BACHER, F., La normalisation de la notation, in *Docimologie et Education*, numéro spécial de la revue « Les Sciences de l'Education », n° 2-3, 1969, 131-156.
- BACHER, F., L'évaluation des résultats scolaires au niveau de l'école moyenne, in *Le Travail humain*, 1965, 28, 219-230.
- BAZIN, R., Les Français d'aujourd'hui et leurs examens, in *Education et Gestion*, 4, 1970, 3-10.
- BLOCK, J., *Mastery Learning*, New York, Holt, Rinehart and Winston, 1970.
- BLOOM, B., HASTING and MADDAUS, *Formative and Summative Evaluation of Student Learning*, New York, Mc Graw-Hill, 1970.
- BONNARDEL, R., Application de la méthode d'analyse factorielle de Thurstone à l'étude de la notation des copies d'examens, in *Le Travail humain*, VIII, 1946, 130-139.
- BRITTON, J., Experimental Marking of English Composition Written by Fifteen-Year-Olds, in *Educational Review*, Birmingham, Vol. 16.1, 1963, 17-23.
- BRITTON, J., MARTIN, N. et ROSEN, H., *Multiple Marking of English Composition, An Account of an Experiment*, London, H.M.S.O., 1966.
- BRUCE, G., *Secondary School Examinations. Facts and Commentary*, Oxford, Pergamon Press, 1969.
- BRUNELLE, L., *Pourquoi des examens ?* Paris, Société des Editions rationalistes, 1968.
- CHOPPIN, B. et PURVES, A., A comparison of open-ended and multiple choice items dealing with literary understanding, in *Research in the Teaching of English*, 3, 1, 1969, 15-24.
- DEMANGEON, M. et LARCEBEAU, S., Une expérience de correction multiple, in *BINOP*, 1958, 14, 131-156.
- Docimologie et Education*, numéro spécial de la revue *Les sciences de l'Education*, 2-3, 1969, 166 p.
- EDGEWORTH, F.V., The Statistics of Examinations, in *Journal of the Royal Stat. Society*, 1888, 51, 599-635.
- Examen des examens*, N° spécial des *Cahiers de pédagogie*, 92, septembre 1970.
- Examinations Bulletins*, Londres, H.M.S.O.
- N° 1. *The Certificate of secondary education : some suggestions for teachers and examiners*, 1963.
- N° 2. *The C.S.E. : Experimental examinations - Mathematics*, 1964.
- N° 3. *The C.S.E. : An introduction to some techniques of examining*, 1964.
- N° 4. *The C.S.E. : An introduction to objective-types examinations*, 1964.

- N° 5. *The C.S.E. : School-based examinations*, 1965.
- N° 6. *The C.S.E. : Experimental examinations : Technical drawing*, 1965.
- N° 7. *The C.S.E. : Experimental examinations - Mathematics 2*, 1965. 1964.
- N° 8. *The C.S.E. : Experimental examinations : Science*, 1965.
- N° 9. *The C.S.E. : Trial examinations : Home economics*, 1966.
- N° 10. *The C.S.E. : Experimental examinations : Music*, 1966.
- N° 11. *The C.S.E. : Trial examinations - Oral English*, 1966.
- N° 12. *Multiple marking of English compositions*, 1966.
- N° 13. *The C.S.E. : Trial examinations : Handicraft*, 1966.
- N° 14. *The C.S.E. : Trial Examinations - Geography*, 1966.
- N° 15. *Teachers' experience of school based examining (English and Physics)*, 1967.
- N° 16. *The C.S.E. : Trial examinations - Written English*, 1967.
- N° 17. *The C.S.E. : Trial examinations - Religious knowledge*, 1967.
- N° 18. *The C.S.E. : The place of the personal topic - History*, 1968.
- N° 19. *C.S.E. : Practical work in science*, 1969.
- N° 20. *C.S.E. : A Group Study Approach to Research and Development*, Londres, Evans-Methuen, 1970.
- FISCHER, H., Wechselwirkungen zwischen Unterrichtszielen, Didaktik und Prüfungen, in *Eidgenössischen Technischen Hochschulen Bulletin* (Zürich), août 1970, 9-14.
- FRENCH, J.W., *Schools of Thought in Judging Excellence in English Themes*, Princeton, E.T.S., 1961.
- HARTOG, P. and RHODES, E.C., *An Examination of Examinations*, London, McMillan, 1936.
- HARTOG, P., *The Marking of English Essays*, London, McMillan, 1941.
- HINTON, E.M., *An Analytical Study of the Qualities of Style and Rhetoric Found in English Compositions*, New York, 1940.
- HOTYAT, F., *Les examens*, Paris, Bourrelier, 1962.
- INGENKAMP, K., *Die Fragwürdigkeit der Zensurengebung*, Weinheim, Beltz, 1971.
- LAUGIER, H. et SCHREIDER, E., Recherche docimologique sur un examen de l'enseignement supérieur, in *Biotypologie*, 1958, 19, n° 2, 61-72.
- LAUWERYS, J.A. et SCANLON, D.G., Ed., *Examinations*, The World Year Book of Education, 1969, London, Evans, 1969.
- LLOYD, W.A., Les examens en Angleterre, in *Revue Française de Pédagogie*, janvier 1968.
- MATHER, D., FRANCE, N. et SARE, G., *The C.S.E., A Handbook for Moderators*, London, Collins, 1965.
- McINTOSH, D., WALKER, D. and McKAY, D., *The Scaling of Teachers' Marks and Estimates*, Edinburgh, Oliver and Boyd, 1962, 2^e éd.
- MONTGOMERY, R.J., *Examinations. An Account of their Evolution as Administrative Devices in England*, Londres, Longmans, 1965.

- JOINT MATRICULATION BOARD
History, Universities of Manchester and Birmingham, 1964.
- OTTER, H.S., *A Functional Language*
- PASSERON, J.C., *Sociologie des Examens*, 6-16.
- PEDLEY, F.H., *A Parent's Guide to Examinations*, 1964.
- PIDGEON, D. et YATES, A., *Examinations*, Londres, Routledge et Kegan Paul, 1957.
- PIERON, H., *Examens et docimologie*
- PIOBETTA, J.B., *Examens et Contrôle*
- REMONDINO, C., Recherche sur les examens, in *Le Travail humain*, 1967, (1st) - 1964 (8st).
- Reports of the Secondary Schools Examinations, 1947, (1st) - 1964 (8st).
- REUCHLIN, M., *L'orientation personnelle*, Conseil de l'Europe, C.C.C., 1964.
- REUCHLIN, M. et BACHER, F., Les examens, in *Revue française de Pédagogie*, 1964.
- ROLLER, S., *L'évaluation du travail individuel et collectif* (Montreux), 1970, 3^e éd.
- ROT, N. et BUJAS, Z., Les distributions des résultats des examens, in *Le Travail humain*, 1959, 22, 1.
- THORNDIKE, R.L., *Marks and Measures*, *Journal of Educational Research*, London, 1927.
- Symposium sur la docimologie, in *Revue de Psychologie* (Rome), 1958, juin 1959.
- TYLER, R., GAGNE, R. et SCRIVAN, M., *Examinations*, AERA Monograph series, Mc Nally, 1967.
- VALENTINE, C.W., *The Reliability of Examinations*, 1932.
- VERNON, P., *Secondary School Examinations*
- WALKER, A.S., *Pupils' School Examinations*
- WENDLER, J., *Standardarbeiten und Examinationsverfahren*, Weinheim, J. Beltz, 1967.
- WISEMAN, S., *The Marking of Examinations*, in *British Journal of Educational Research*, 1961.
- WISEMAN, S., *Examinations and Assessment*, London, Methuen Press, 1961.
- WRIGLEY, J., *The Relative Efficiency of Examinations*, in *Journal of Educational Psychology*, 25, 1955, 10.
- YATES, A. and PIDGEON, D., *Examinations*, N.F.E.R., 1957.

nations, 1965.
nations : Technical drawing, 1965.
nations - Mathematics 2, 1965.
nations : Science, 1965.
: Home economics, 1966.
nations : Music, 1966.
Oral English, 1966.
ositions, 1966.
Handicraft, 1966.
Geography, 1966.
based examining (English and
Written English, 1967.
- Religious knowledge, 1967.
ersonal topic - History, 1968.
1969.
h to Research and Development,
Unterrichtszielen, Didaktik und
chen Hochschulen Bulletin (Zü-
g Excellence in English Themes,
nation of Examinations, London,
, London, McMillan, 1941.
Qualities of Style and Rhetoric
, 1940.
1962.
nsurengung, Weinheim, Beltz,
ocimologique sur un examen de
, 1958, 19, n° 2, 61-72.
aminations, The World Year Book
Revue Française de Pédagogie,
C.S.E., A Handbook for Modera-
The Scaling of Teachers' Marks
, 1962, 2^e éd.
unt of their Evolution as Admi-
ngmans, 1965.

- JOINT MATRICULATION BOARD, *The Marking of Scripts in Advanced Level History*, Universities of Manchester, Liverpool, Leeds, Sheffield and Birmingham, 1964.
- OTTER, H.S., *A Functional Language Examination*, Oxford Univ. Press, 1968.
- PASSERON, J.C., Sociologie des examens, in *Education et Gestion*, 1970, 2, 6-16.
- PEDLEY, F.H., *A Parent's Guide to Examinations*, Oxford, Pergamon Press, 1964.
- PIDGEON, D. et YATES, A., *An Introduction to Educational Measurement*, Londres, Routledge et Kegan Paul, 1968.
- PIERON, H., *Examens et docimologie*, Paris, P.U.F., 1963.
- PIOBETTA, J.B., *Examens et Concours*, Paris, P.U.F., 1943.
- REMONDINO, C., Recherche sur les systèmes numériques d'évaluation scolaire, in *Le Travail humain*, 1965, 18, 3-4, 263-265.
- Reports of the Secondary School Examinations Council*, Londres, H.M.S.O., 1947, (1st) - 1964 (8st).
- REUHLIN, M., *L'orientation pendant la période scolaire*, Strasbourg, Conseil de l'Europe, C.C.C., 1964.
- REUHLIN, M. et BACHER, F., L'appréciation des élèves par leurs professeurs, in *Revue française de Pédagogie*, 1968, 2, 19-25.
- ROLLER, S., L'évaluation du travail pédagogique, in *Educateur et bulletin corporatif* (Montreux), 1970, 36, 694-696.
- ROT, N. et BUJAS, Z., Les distributions de notes scolaires comparées aux distributions des résultats obtenus dans les tests de connaissances, in *Le Travail humain*, 1959, 22, 19-26.
- THORNDIKE, R.L., Marks and Marking Systems, in R.L. EBEL, *Encyclopedia of Educational Research*, Londres, McMillan, 1969, pp. 759-766.
- Symposium sur la docimologie, XIII^e congrès de l'Association Internationale de Psychologie (Rome, 1958), in *Le Travail humain*, XXII, 1-2, janvier-juin 1959.
- TYLER, R., GAGNE, R. et SCRIVEN, N., *Perspectives of Curriculum Evaluation*, AERA Monograph series on curriculum eval., N° 1, Chicago, Rand Mc Nally, 1967.
- VALENTINE, C.W., *The Reliability of Examinations*, University of London Press, 1932.
- VERNON, P., *Secondary School Selection*, Londres, Methuen, 1957.
- WALKER, A.S., *Pupils' School Records*, Newnes, Educ. Publ., 1955.
- WENDLER, J., *Standardarbeiten, Verfahren zur Objektivierung der Notenggebung*, Weinheim, J. Beltz, 1969.
- WISEMAN, S., The Marking of English Composition in Grammar School Selection, in *British Journal of Educational Psychology*, XIX, 1949, 200-209.
- WISEMAN, S., *Examinations and English Education*, Manchester, University Press, 1961.
- WRIGLEY, J., The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools, in *British Journal of Educational Psychology*, 25, 1955, 107-116.
- YATES, A. and PIDGEON, D., *Admission to Grammar Schools*, London N.F.E.R., 1957.

TABLE ANALYTIQUE

INTRODUCTION	7
------------------------	---

PREMIERE PARTIE

DEFINITIONS

I. Docimologie, docimastique et doxologie	13
II. Examens et concours :	
Observation et évaluation continues	14
Examens internes et examens externes	15
III. Notes et scores	16

DEUXIEME PARTIE

L'ACCUSATION ET LA DEFENSE

CHAPITRE 1. — Critique des examens	19
1. Corps étrangers dans l'éducation, au service d'une pédagogie dépassée	19
2. Anxiété et <i>stress</i>	20
3. Inégalité - injustice	21
4. L'échec, générateur d'échecs	25
5. Rupture entre enseignement et examens	26
6. Désaccord entre correcteurs	27
a) Composition française	
b) Mathématiques	
c) Médecine	
d) Divers	
e) Combien de correcteurs pour stabiliser la note ?	
7. Infidélité d'un même correcteur	
Un schéma pour continuer la recherche	32
8. Stéréotypes et effets de halo	33

9. Manque de validité	
10. Un instrument d'immobilisme :	
a) Effets irréversibles de la	
b) Les examens ne sont pas	
11. Faiblesse de beaucoup d'exp	
12. Autres critiques	

CHAPITRE 2. — Défense de la n

1. La mesure rigoureuse est peut	
2. Les maîtres jugent bien les é	
3. Validité limitée mais réelle d	
4. S'endurcir pour la vie	
5. Se situer par rapport aux a	
6. Large synthèse et intégration	
7. L'examen externe contrôle l	
8. L'examen externe, <i>feedback</i>	

TRC

CONSTRUC

Les grandes phases - Vue d'ens

CHAPITRE 1. — L'objet et les ob

I. L'objet	
A. Le pronostic	
1. Tests de maturité spéc	
2. Vérification des conna	
3. Essai	
B. L'inventaire	
C. Le diagnostic	
II. Les objectifs	
A. Les objectifs généraux	
1. Les objectifs cognitifs	

7

13

14

15

16

19

ice d'une pédagogie dé- 19

20

21

25

26

27

la note ?

32

33

9. Manque de validité	36
10. Un instrument d'immobilisme social	37
a) Effets irréversibles de la certification scolaire	37
b) Les examens ne sont pas neutres	38
11. Faiblesse de beaucoup d'expériences docimologiques	40
12. Autres critiques	40

CHAPITRE 2. — Défense de la note subjective et de l'examen 42

1. La mesure rigoureuse est peut-être impossible	42
2. Les maîtres jugent bien les élèves	44
3. Validité limitée mais réelle des examens traditionnels	46
4. S'endurcir pour la vie	47
5. Se situer par rapport aux autres	47
6. Large synthèse et intégration des connaissances	47
7. L'examen externe contrôle le professeur	47
8. L'examen externe, <i>feedback</i> pour le professeur	48

TROISIEME PARTIE

CONSTRUCTION DE L'EXAMEN

Les grandes phases - Vue d'ensemble	51
---	----

CHAPITRE 1. — L'objet et les objectifs 52

I. L'objet 52

A. Le pronostic 52

 1. Tests de maturité spécifique (*readiness*) 53

 2. Vérification des connaissances-clés ou notions critiques 53

 3. Essai 54

B. L'inventaire 54

C. Le diagnostic 54

II. Les objectifs 56

A. Les objectifs généraux 57

 1. Les objectifs cognitifs 57

a) La taxonomie de Bloom	58
b) Le modèle de J.-P. Guilford	60
2. Domaine affectif	64
B. Les objectifs spéciaux	66
Illustration	68
CHAPITRE 2. — Rédaction des questions	73
I. Observations générales	73
A. Des questions compréhensibles	74
B. Tenir compte du niveau d'information	74
C. Essayer ou prétester les questions	74
D. Calcul de la facilité des questions	75
E. Calcul de l'efficacité. Pouvoir discriminatif	75
1. Méthode simple	75
2. Méthode plus fine	76
II. Réponses ouvertes ou fermées ?	79
A. Réponses ouvertes	79
B. Réponses fermées - Questions à choix multiple	81
1. Utilité	81
2. Constituer une provision de questions	81
3. Exploiter la gamme des possibilités logiques	82
a) Question à complément simple	82
b) Association simple	83
c) Association composée	83
d) Association à terme exclu	84
e) Analyse de relations de cause à effet	84
f) Analyse d'observations	84
g) Comparaisons quantitatives	86
h) Compléments groupés	87
4. Calcul de l'efficacité des distracteurs	87
5. Critiques et réfutation partielle	88
a) Une objectivité trompeuse	88

b) Choix « corrects » con
c) Un jeu de hasard
d) Acrobatie mentale
e) Inconvénients incertai
C. En guise de conclusion : u

CHAPITRE 3. — La notation
I. Un préambule indispensable :
A. La courbe de Gauss, image
B. La courbe de Gauss, image
C. L'écart type ou sigma, ind
1. Signification
2. Estimation rapide
D. La concentration des résultats
E. Courbe de Gauss voulue pa
F. Comment savoir si une dist
II. La notation subjective : l'éch
A. Introduction
B. Nature et faiblesses des éch
C. Utilité
D. Construction
1. Combien de degrés ?
2. Définir l'objet de l'évalua
E. Utilisation
1. Combien d'élèves par éc
— Elève comparé à lui-m
— Elèves comparés entre
2. Lutter contre la contamina
F. Comment synthétiser les éval

	58
	60
	64
	66
	68
	73
	73
	74
	74
	74
	75
matif	75
	75
	76
	79
	79
ix multiple	81
	81
s	81
ogiques	82
	82
	83
	83
	84
et	84
	84
	86
	87
	87
	88
	88

b) Choix « corrects » contestables	89
c) Un jeu de hasard	89
d) Acrobatie mentale	90
e) Inconvénients incertains	91
C. En guise de conclusion : un compromis	91
CHAPITRE 3. — La notation	93
I. Un préambule indispensable : la courbe de Gauss	93
A. La courbe de Gauss, image de la probabilité	93
B. La courbe de Gauss, image de l'enseignement non individualisé	95
C. L'écart type ou sigma, indice précieux	96
1. Signification	96
2. Estimation rapide	97
D. La concentration des résultats autour de la moyenne	100
E. Courbe de Gauss voulue par les maîtres	101
F. Comment savoir si une distribution est normale ?	101
II. La notation subjective : l'échelle d'évaluation	105
A. Introduction	105
B. Nature et faiblesses des échelles d'évaluation	106
C. Utilité	107
D. Construction	107
1. Combien de degrés ?	108
2. Définir l'objet de l'évaluation	108
E. Utilisation	111
1. Combien d'élèves par échelon ?	111
— Elève comparé à lui-même	111
— Elèves comparés entre eux	113
2. Lutter contre la contamination et la tendance centrale	114
F. Comment synthétiser les évaluations ?	115

G. Un cas particulier : la notation de la composition française	118
1. Quatre méthodes d'évaluation	118
a) La méthode de l'impression générale	118
b) Les échelles de spécimens	120
c) La méthode analytique	121
Quelles qualités observer ?	122
Un exemple	124
d) La méthode des comptages de fréquences	127
2. Plusieurs sujets au choix ?	128
3. Conclusion	129
III. La notation objective	130
IV. L'étalonnage ou mesure de la position relative	130
A. Le centilage	131
B. Les notes standard ou notes Z	134
C. L'échelle à cinq classes	136
D. L'échelle à neuf classes (stanines)	137
CHAPITRE 4. — Contrôle de la fidélité de l'examen	138
1. Eviter toute ambiguïté dans les questions	138
2. Des questions en nombre suffisant	139
3. Un contrôle mathématique	139
a) La méthode pairs-impairs	139
b) Deux formes parallèles	140
4. Répétition de la notation	140
CHAPITRE 5. — Contrôle de la validité	141
I. La validité du contenu	141
II. La validité prédictive	143

QUATRIEME PARTIE

LES PROCEDURES DE MODERATION

CHAPITRE 1. — Position du problème	147
1. Définition	147

2. Modérer n'est pas capor	
3. Modération volontaire ou	
4. La modération commence	
5. Pas de comparabilité sa	
6. Peut-on se fier aux tests	

CHAPITRE 2. — Quelques systè

I. La formule la plus libérale	
II. Système imposé de modérati connaissances	
III. Modération volontaire par app	
IV. Un système simple au service	
V. En Angleterre, un système de	
A. Préliminaires	
B. Les professeurs notent les	
C. Correction par les modéra	
1° Contrôle de la sévérité	
2° Contrôle de la discrimi	
3° Contrôle de la confor	
D. Correction des échantillon	
E. Comment ajuster des notes	
1° Ajustement du médian	
2° Ajustement du médian	
VI. La note de fin d'année	
Travail de l'année + travaux	

CINC

LE MYTHE DE

CHAPITRE 1. — Evolution de la c	
I. La courbe des aptitudes	
II. La courbe des connaissances	

a composition française	118
.	118
erale	118
.	120
.	121
.	122
.	124
fréquences	127
.	128
.	129
.	130
relative	130
.	131
.	134
.	136
.	137
ramen	138
os	138
.	139
.	139
.	139
.	140
.	140
.	141
.	141
.	143

IE

MODERATION

.	147
.	147

2. Modérer n'est pas caporaliser	148
3. Modération volontaire ou imposée ?	149
4. La modération commence au début de l'année scolaire	149
5. Pas de comparabilité sans fidélité élevée	150
6. Peut-on se fier aux tests ?	151
CHAPITRE 2. — Quelques systèmes de modération	153
I. La formule la plus libérale : le système suédois	153
II. Système imposé de modération par branche à partir d'un test de connaissances	156
III. Modération volontaire par appel à une banque de questions	156
IV. Un système simple au service de la sélection	157
V. En Angleterre, un système de modération complet	160
A. Préliminaires	160
B. Les professeurs notent les examens	161
C. Correction par les modérateurs	161
1° Contrôle de la sévérité	163
2° Contrôle de la discrimination	164
3° Contrôle de la conformité	165
D. Correction des échantillons restants et contrôles	165
E. Comment ajuster des notes discordantes ?	168
1° Ajustement du médian	168
2° Ajustement du médian et de l'écart type	168
VI. La note de fin d'année.	
Travail de l'année + travaux pratiques + test	171

CINQUIEME PARTIE

LE MYTHE DE LA COURBE DE GAUSS

CHAPITRE 1. — Evolution de la courbe des connaissances	180
I. La courbe des aptitudes	180
II. La courbe des connaissances	181

CHAPITRE 2. — Une pédagogie de la courbe en j	184
CHAPITRE 3. — La théorie de l'évaluation formative	186
I. Jalonner l'ascension du savoir	187
II. Guider l'élève	188
III. Limites pratiques du système	189
CONCLUSIONS ET RECOMMANDATIONS	191

ANNEXES

I. Etude comparée d'une question d'examen présentée selon la méthode traditionnelle et selon la méthode par questions à choix multiple	199
— Méthode traditionnelle	199
— Méthode par choix multiple	200
III. Composition de langue maternelle - Angleterre, G.C.O., 1967 - Fin du secondaire	205
— Questions	205
Epreuve I	205
Epreuve II	206
— Consignes pour la correction	208
Epreuve I	208
Epreuve II	210
III. Exemple d'enseignement semi-individualisé.	
— La New Trier Township High School	214
— L'individualisation des programmes	215
— L'individualisation de l'enseignement	218
— L'étude de chaque élève avant son entrée à New Trier	223
— Les conseillers pédagogiques	225

TAB

INTRODUCTION	
PREMIERE PARTIE — <i>Délimitation</i>	
DEUXIEME PARTIE — <i>L'accusation</i>	
Chapitre 1 : Critique des	
Chapitre 2 : Défense de la	
TROISIEME PARTIE — <i>Construction</i>	
Chapitre 1 : L'objet et les	
I. L'objet	
II. Les objectifs	
Chapitre 2 : Rédaction des	
I. Observations générales	
II. Réponses ouvertes	
Chapitre 3 : La notation	
I. Un préambule introductif	
II. La notation subjective	
III. La notation objective	
IV. L'étalement des notes	
Chapitre 4 : Contrôle de	
Chapitre 5 : Contrôle de	

...rbe en j	184
...formative	186
...	187
...	188
...	189
...	191
...	199
...	199
...	200
...terre, G.C.O., 1967 - Fin	205
...	205
...	205
...	206
...	208
...	208
...	210
...	214
...	215
...	218
... à New Trier	223
...	225

TABLE DES MATIERES

INTRODUCTION	7
PREMIERE PARTIE — <i>Définitions</i>	11
DEUXIEME PARTIE — <i>L'accusation et la défense</i>	17
Chapitre 1 : Critique des examens	19
Chapitre 2 : Défense de la note subjective et des examens	42
TROISIEME PARTIE — <i>Construction de l'examen</i>	49
Chapitre 1 : L'objet et les objectifs	52
I. L'objet	52
II. Les objectifs	56
Chapitre 2 : Rédaction des questions	73
I. Observations générales	73
II. Réponses ouvertes ou fermées ?	79
Chapitre 3 : La notation	93
I. Un préambule indispensable. La courbe de Gauss	93
II. La notation subjective. Echelle d'évaluation	105
III. La notation objective	130
IV. L'étalonnage ou mesure de la position relative	130
Chapitre 4 : Contrôle de la fidélité	138
Chapitre 5 : Contrôle de la validité	141

QUATRIEME PARTIE — <i>Les procédures de modération</i>	145
Chapitre 1 : Position du problème	147
Chapitre 2 : Quelques systèmes de modération	153
I. La formule suédoise, la plus libérale - Système de modération par branche à partir d'un test de connaissances	153
II. Système imposé sur la même base.	156
III. Modération volontaire, par appel à une banque d'items	156
IV. Un système simple de modération globale au service de la sélection	157
V. En Angleterre. Un système de modération complet	160
VI. Travail de l'année + travaux pratiques + test. Quel grade attribuer en fin d'année ?	171
CINQUIEME PARTIE — <i>Le mythe de la courbe de Gauss</i>	177
Chapitre 1 : Evolution de la courbe des connaissances	180
Chapitre 2 : Une pédagogie de la courbe en J	184
Chapitre 3 : La théorie de l'évaluation formative	186
CONCLUSION ET RECOMMANDATIONS	191
ANNEXES	197
I. Etude comparée d'une question d'examen	199
II. Exemple de questions pour une composition en langue maternelle	205
III. Exemple d'enseignement semi-individualisé	214
BIBLIOGRAPHIE SUCCINCTE	227
TABLE ANALYTIQUE	230

ACHEVE
LE 10

modération	145
.	147
modération	153
s libérale - Système de partir d'un test de connais-	153
base.	156
rel à une banque d'items	156
ration globale au service	157
modération complet	160
pratiques + test. Quel ée ?	171
brbe de Gauss	177
connaissances	180
e en J	184
formative	186
.	191
.	197
d'examen	199
composition en langue	205
-individualisé	214
.	227
.	230

ACHEVE D'IMPRIMER
LE 10 MARS 1972.