

CONSTRUIRE DES ECHELLES D'EVALUATION DESCRIPTIVES

par

R. DE BAL
G. DE LANDSHEERE
J. PAQUAY-BECKERS



BRUXELLES
MINISTÈRE DE L'ÉDUCATION NATIONALE
ET DE LA CULTURE FRANÇAISE
DIRECTION GÉNÉRALE DE L'ORGANISATION DES ÉTUDES
1976

TABLE DES MATIÈRES

	Pages
INTRODUCTION	9
PREMIÈRE PARTIE	
L'ÉVALUATION : QUELQUES CONDITIONS NÉCESSAIRES	
◊ A. L'évaluation doit être <i>valide</i>	21
◊ B. L'évaluation doit être <i>objective</i>	25
. Pour informer l'élève	26
. Pour informer l'enseignant	26
C. L'évaluation doit être <i>formative</i>	26
. Porter sur la performance et non sur la personne	27
. Etre analytique	27
. Les objectifs doivent être opérationnalisés	28
DEUXIÈME PARTIE	
ENQUÊTE SUR L'ÉVALUATION	
A. Contexte	33
B. Objectifs	35
C. Population et disciplines concernées	35
D. Questionnaire	36
E. Dépouillement	36
F. Résultats	36
1. Modes de contrôle	36
2. Fréquence des contrôles	42
3. Synthèse des évaluations en fin de période	48
4. Coordination des contrôles	51
5. Normes	56
6. Difficultés rencontrées	60
7. Remarques complémentaires faites par les professeurs	63
G. Conclusions	65

TROISIÈME PARTIE

CONSTRUCTION D'ÉCHELLES D'ÉVALUATION DESCRIPTIVES

CHAPITRE 1 : <i>Les échelles d'évaluation</i>	69
A. Deux grands types de mesures	69
B. Espèces d'échelles d'évaluation	71
C. Quand utiliser des échelles d'évaluation ?	73
D. Phénomènes menaçant l'objectivité et la validité des évaluations	77
E. Nécessité de définitions opérationnelles	80
CHAPITRE 2 : <i>Étapes et conditions de la recherche</i>	82
A. Étapes	82
B. Conditions	82
CHAPITRE 3 : <i>Méthodologie de la construction d'échelles d'évaluation descriptives</i>	85
A. Construction initiale de l'instrument	85
I. Première étape : Sélection et définition des critères d'évaluation	85
a) Un petit nombre de critères	85
b) Comment définir les critères	88
II. Deuxième étape : Sélection et définition des aspects de chaque critère	89
III. Troisième étape : Détermination des niveaux d'exigence pour chaque aspect	92
IV. Quatrième étape : Description des échelons	97
V. Cinquième étape : Présentation de l'échelle et démarches pour son utilisation	99
B. Remaniement de l'instrument	102
I. Recherche des vices de construction et d'utilisation	102
II. Comment réduire les discordances entre juges ?	106
CHAPITRE 4 : <i>Résultats expérimentaux</i>	108
I. Première expérimentation	108
a) Schéma expérimental	108
b) Échantillon	110
II. Seconde expérimentation	111
III. Hypothèses à vérifier	111
IV. Dépouillement des données	112
V. Résultats	113

<i>Hypothèse 1 : L'échelle d'évaluation est objective.</i>
Comment vérifier l'hypothèse ?
1° Concordance des juges
2° Étude de la corrélation
3° Étude de l'influence des juges
4° Étude de l'influence des items
<i>Hypothèse 2 : L'échelle d'évaluation est fiable.</i>
Comment vérifier l'hypothèse ?
1° Corrélation des juges
2° Influence des juges
<i>Hypothèse 3 : L'échelle d'évaluation est valide.</i>
Comment vérifier l'hypothèse ?
1° Le nombre de juges
2° Les professeurs jugent-ils les objectifs de la même manière ?
3° Les commentaires sont-ils facilement identifiables ?
4° Les professeurs ont-ils des attitudes positives ou négatives ?

CONCLUSION

ANNEXES

A. Échelles d'évaluation
1. Rédaction : comment rédiger un questionnaire secondaire ?
2. Résumé et critique
3. Sciences : compte rendu de procédures de validation
4. Sciences : observation de l'utilisation d'une méthode d'évaluation
B. Résultats détaillés

	Pages
ON DESCRIPTIVES	
.....	69
.....	69
.....	71
.....	73
.....	77
.....	80
.....	82
.....	82
.....	82
.....	85
.....	85
.....	85
.....	88
.....	89
.....	92
.....	97
.....	99
.....	102
.....	102
.....	106
.....	108
.....	108
.....	108
.....	110
.....	111
.....	111
.....	112
.....	113

	Pages
<i>Hypothèse 1</i> : L'échelle descriptive permet une évaluation plus objective.	113
Comment vérifier cette hypothèse ?	113
1° Concordance des notes partielles considérées isolément	114
2° Etude de la concordance entre les profils de notes	116
3° Etude de l'influence des facteurs externes	117
4° Etude de l'influence des facteurs internes	119
<i>Hypothèse 2</i> : L'échelle descriptive permet une évaluation plus valide	120
Comment vérifier cette hypothèse ?	120
1° Corrélation entre la note globale et chacune des appréciations analytiques	120
2° Influence respective des notes analytiques sur la note globale	121
<i>Hypothèse 3</i> : L'échelle descriptive permet une évaluation mieux utilisable par le professeur et par l'élève	122
Comment vérifier cette hypothèse ?	122
1° Le nombre de justifications diffère-t-il d'un groupe à l'autre ?	122
2° Les professeurs utilisent-ils des justifications en rapport avec les objectifs du travail proposé ?	122
3° Les commentaires destinés à l'élève concernent-ils des critères facilement identifiables par ce dernier ?	123
4° Les professeurs relèvent-ils, dans leurs commentaires, les aspects positifs ou négatifs du travail ?	123
CONCLUSION	124
ANNEXES	127
A. Echelles d'évaluation.	129
1. Rédaction : complètement d'un récit mutilé (1 ^{er} cycle de l'enseignement secondaire rénové)	131
2. Résumé et critique d'un texte (2 ^e cycle)	153
3. Sciences : compte rendu d'une expérience, interprétation, invention de procédures de mesure (1 ^{er} cycle)	171
4. Sciences : observation d'une expérience, interprétation, invention d'une méthode d'analyse (2 ^e cycle)	189
B. Résultats détaillés	209

INTRODUCTION

La rénovation de l'enseignement secondaire belge, bientôt étendue à l'enseignement primaire, a été immédiatement accompagnée d'une réforme des méthodes d'évaluation. Cette réforme a pour principaux objectifs :

- d'accorder la prépondérance à la fonction éducative de l'évaluation, trop longtemps réalisée dans le passé à des fins surtout sélectives;
- d'élargir systématiquement la palette évaluative, non seulement pour couvrir tout le domaine cognitif, mais aussi pour faire une large place au domaine affectif;
- d'introduire dans la vie scolaire des méthodes et des techniques d'évaluation présentant plus de garanties scientifiques que les procédures habituelles.

Une rénovation aussi profonde du système éducatif ne peut pas s'opérer du jour au lendemain. D'abord, elle concerne des dizaines de milliers d'enseignants et des centaines de milliers d'élèves. Ensuite, elle met en cause, chez les maîtres, des habitudes, des attitudes, des croyances acquises au cours de leurs études, puis renforcées et approfondies pendant une carrière professionnelle parfois longue; la réforme exige d'ailleurs un vaste effort de recyclage qui s'étendra sur plusieurs années. Enfin, il ne suffit pas de décider de changer de méthodes et d'instruments d'évaluation; il faut encore que les nouveaux instruments existent et que ceux à qui ils sont destinés les acceptent et sachent s'en servir correctement. Disons tout de suite que, pour remplir ces deux conditions, il importe que les enseignants participent directement aux décisions de principe et à la construction des nouveaux instruments.

Assurément, l'évaluation ne se réduit pas à un problème d'outils. Toute l'intelligence, toute la sensibilité de l'éducateur interviennent dans les jugements qu'il porte et les appréciations qu'il formule au cours des interactions multiformes entre le maître et ses élèves. Quelle que soit la banalité de la constatation, rappeler que la pédagogie est autant art que science s'impose lorsqu'on traite d'évaluation en matière d'éducation.

Cela étant dit, quelles sont les méthodes fondamentales de l'évaluation? Elle est analytique ou globale, objective ou subjective, mathématique ou affective, ces deux aspects ne s'excluant pas nécessairement, mais venant, au contraire, se compléter d'heureuse façon : on peut compter le nombre de fautes d'orthographe commises par un élève, puis

émettre un jugement relatif (performance meilleure que la précédente, performance de bonne qualité vu l'âge du sujet,...).

Dans l'exemple que nous venons de proposer, le *calcul* d'un score est suivi d'une appréciation globale. Inversement, une performance dont la qualité est d'abord jugée globalement, peut ensuite se voir affecter une note chiffrée pour établir une hiérarchie entre personnes ou entre productions successives d'un même individu.

Mais ces observations ne nous ont pas encore conduit au fond du problème. Pour éviter tout exemple scolaire qui pourrait être affectivement chargé, plaçons-nous un instant sur le terrain du sport et comparons la façon d'évaluer une course de vitesse à patins et une épreuve de patinage artistique. Toutes deux sont susceptibles d'entraîner des appréciations affectives, mais seule la course se prête à la mesure exacte, objective. Pourquoi ? Parce que, dans la course, un seul critère, appréciable par simple comptage, importe : le temps. On compte (avec la précision que l'on sait) le nombre de secondes, de minutes,... On mesure, dans ce cas, UNE SEULE MODALITÉ du comportement. Pour la facilité, on parle ici d'une PERFORMANCE SIMPLE (ce qui ne signifie nullement qu'elle est aisée ou qu'elle résulte d'un processus sans complexité !).

Par contre, il est impossible d'ainsi réduire le patinage artistique. On peut encore imaginer des dispositifs d'appréciation du rythme, recourir à un inventaire de comportements pour juger de l'originalité des figures, mais qu'en sera-t-il lorsqu'il faudra évaluer le caractère plus ou moins gracieux des évolutions, la beauté de la performance, la fluidité des enchaînements... ? Les juges sportifs font sagement la distinction entre les notes techniques (susceptibles d'évaluation par décompte de points) et les notes artistiques qui, en dernière analyse, deviennent globales. Ici, on tient compte d'un FAISCEAU DE MODALITÉS. La PERFORMANCE est dite COMPLEXE.

La complexité varie naturellement selon les cas. Mais il ne faut guère de facteurs susceptibles d'interactions pour arriver à des tous non réductibles : l'ensemble a pris le pas sur les comportements isolés ⁽¹⁾.

(1) On remarquera aussi que, pour mesurer le temps d'une course, un seul chronomètre électronique suffit, alors que, lors d'un championnat, la note artistique, en patinage, résulte de la moyenne de onze évaluations formulées simultanément.

Nous reviendrons, par
du présent ouvrage, ma
une première fois le prob
l'école a besoin d'instru
(tests objectifs) et d'éc
plexes.

Or, au moment où la
fondamental, ses écoles r
sances et l'on peut consi
perfectionnées n'existaie
urgent besoin de tests p
d'échelles d'appréciation

On sait ce qu'il advint
défaut; l'usage de simpl
imposé. Accuser les aut
pas attendu que les ins
d'engager la réforme sera
celle que nous travers
peuvent être différées, e
comprendre pourquoi, e
jusqu'à présent, réserv
fique, il n'en reste pas r
de construire ce qui nou

Très vite, on a senti q
à cinq ou sept degrés, t
général, une utile réflex
encore loin d'une évalu
prendre la situation, nou
de tous les professeurs
tionnant dans l'enseigne
à la fois combien grand
et combien maigres étaie
par rapport au système

De toute évidence, il
note A pour faire naît
L'important, comme il

leure que la précédente,
et...).

ser, le *calcul* d'un score
t, une performance dont
ensuite se voir affecter
entre personnes ou entre

pre conduit au fond du
i pourrait être affective-
rain du sport et compa-
à patins et une épreuve
ceptibles d'entraîner des
prête à la mesure exacte,
s, un seul critère, appré-
ps. On compte (avec la
e minutes,... On mesure,
comportement. Pour la
MPLE (ce qui ne signifie
e d'un processus sans

le patinage artistique.
appréciation du rythme,
ur juger de l'originalité
ra évaluer le caractère
uté de la performance,
ortifs font sagement la
tibles d'évaluation par
ui, en dernière analyse,
AISCEAU DE MODA-
EXE.

s. Mais il ne faut guère
rriver à des tous non
mportements isolés (1).

une course, un seul chrono-
nmat, la note artistique, en
rmulées simultanément.

Nous reviendrons, par d'autres voies, sur ces notions dans le corps du présent ouvrage, mais, d'emblée, il a semblé nécessaire de clarifier une première fois le problème : pour accomplir son travail d'éducation, l'école a besoin d'instruments de mesure de comportements simples (tests objectifs) et d'échelles d'évaluation de comportements complexes.

Or, au moment où la Belgique a décidé de rénover son enseignement fondamental, ses écoles ne disposaient que de très rares tests de connaissances et l'on peut considérer que des échelles d'évaluation quelque peu perfectionnées n'existaient pratiquement pas. Pourtant, on ressent un urgent besoin de tests pronostiques et diagnostiques, et, plus encore, d'échelles d'appréciation dont l'usage pourrait être presque quotidien.

On sait ce qu'il advint : les tests font toujours presque intégralement défaut; l'usage de simples échelles verbales à cinq ou sept degrés a été imposé. Accuser les autorités scolaires d'impéritie parce qu'elles n'ont pas attendu que les instruments nécessaires soient disponibles avant d'engager la réforme serait injuste. Dans une crise de civilisation comme celle que nous traversons, certaines mutations institutionnelles ne peuvent être différées, et même s'il n'est pas sans intérêt d'essayer de comprendre pourquoi, dans notre pays, la formation des maîtres a, jusqu'à présent, réservé une place insuffisante à la pédagogie scientifique, il n'en reste pas moins qu'un des problèmes les plus urgents est de construire ce qui nous manque.

Très vite, on a senti que l'adoption des échelles d'évaluation verbales à cinq ou sept degrés, telles qu'elles étaient proposées, provoquait, en général, une utile réflexion chez les enseignants, mais que l'on était encore loin d'une évaluation rigoureuse. Pour essayer de mieux comprendre la situation, nous avons procédé, dès 1971, à une enquête auprès de tous les professeurs de langue maternelle et de mathématique fonctionnant dans l'enseignement secondaire rénové. Cette enquête montre à la fois combien grande était la bonne volonté des maîtres concernés et combien maigres étaient les progrès apportés par les nouvelles échelles par rapport au système ancien.

De toute évidence, il ne suffit pas de remplacer la note 8/10 par la note A pour faire naître une plus grande rigueur de l'appréciation ! L'important, comme il sera montré dans le présent ouvrage, c'est de

définir de façon aussi univoque que possible les conditions dans lesquelles telle note sera attribuée.

Ainsi est apparue l'urgente nécessité de perfectionner le système initialement proposé par les réformateurs (qui, rendons-leur cette justice, furent les premiers à remettre leurs propositions en question, ce qui nous change heureusement de la vieille relation autoritaire entre décideurs et exécutants).

A leur demande, nous avons tenté de répondre à plusieurs questions :

- Est-il possible de construire des échelles d'évaluation garantissant une fidélité élevée sinon parfaite ?
- Peut-on déceler des types d'exercices fréquents dans la pratique scolaire et construire, pour les évaluer, des échelles susceptibles d'être régulièrement utilisées ou facilement imitées ?
- Quelle quantité de travail représente la construction d'une batterie d'échelles considérées comme satisfaisantes ?
- Une action particulière doit-elle être entreprise pour que les échelles construites soient utilisées correctement ou suffit-il de les mettre simplement à la disposition des écoles ?

Ainsi s'explique la structure de ce livre. Après avoir rappelé les notions de docimologie nécessaires, on présente les principales conclusions se dégageant de l'enquête préliminaire. On passe ensuite à la construction des échelles d'évaluation, en commençant par en expliquer la théorie.

Les domaines pour lesquels des batteries d'échelles ont été construites sont la langue maternelle (exercice d'expression écrite et résumé de texte suivi de critique) et les sciences naturelles (compte rendu d'expérience, méthode de mesure.,,,).

Les critères de l'évaluation ont été choisis en se référant aux résultats de recherches expérimentales, là où il en existait, et à l'avis d'enseignants expérimentés. C'est aussi en coopération avec des commissions d'enseignants et d'inspecteurs que la construction a été réalisée et que les instruments ont été essayés sur le terrain.

La recherche, dont on va lire le compte rendu et les résultats, tente de montrer jusqu'à quel degré de précision on peut arriver, en matière

d'évaluation de perform
mum de connaissances e
pas à la portée de l'e
d'éducateur). EN PR
MONTRER LES LIM
rigueur, plus de justice
mais on n'arrivera que
Car, pour porter les ju
beauté ou le bien, les b

Mais, ces limites ét
supplémentaires se dég

1. Il est possible de co
des exercices réputé
2. La précision des éch
entre des enseignants
3. Toutefois, pour arri
sinon souvent, accep
contraindre par les d
De nouvelles habitu
respecter rigoureuse
Une première exper
sans ce changement
perdre une bonne p
ment construites p
maximum, l'utilisati
entendons que l'on
4. La mise au point c
considérable. Incom
si dévoués soient-ils
constitue pour eux c

Ainsi se confirme, un
boration permanente
les groupes d'enseignant

Par ailleurs, si des ef
par disposer d'un vas

les conditions dans les-

fectionner le système ini-
endons-leur cette justice,
s en question, ce qui nous
bitaire entre décideurs et

re à plusieurs questions :
d'évaluation garantissant

quents dans la pratique
des échelles susceptibles
t imitées ?

nstruction d'une batterie
?

rise pour que les échelles
ou suffit-il de les mettre

Après avoir rappelé les
te les principales conclu-
. On passe ensuite à la
mençant par en expliquer

chelles ont été construites
a écrite et résumé de texte
mpte rendu d'expérience,

u se référant aux résultats
, et à l'avis d'enseignants
c des commissions d'en-
a été réalisée et que les

u et les résultats, tente de
peut arriver, en matière

d'évaluation de performances complexes, en mettant en œuvre le maxi-
mum de connaissances et de moyens disponibles (ce qui n'est évidemment
pas à la portée de l'enseignant qui fait quotidiennement son métier
d'éducateur). EN PROCÉDANT AINSI, ON A AUSSI VOULU
MONTRER LES LIMITES DU SYSTÈME. On peut assurer plus de
rigueur, plus de justice dans l'évaluation des performances complexes,
mais on n'arrivera que très exceptionnellement à la parfaite unanimité.
Car, pour porter les jugements de valeur, qu'il s'agisse d'apprécier la
beauté ou le bien, les balances n'existent pas.

Mais, ces limites étant reconnues et acceptées, quatre conclusions
supplémentaires se dégagent :

1. Il est possible de construire des échelles d'évaluation précises pour
des exercices réputés difficilement appréciables de façon objective.
2. La précision des échelles conduit à un degré de concordance élevée
entre des enseignants évaluant indépendamment.
3. Toutefois, pour arriver à un accord, les enseignants doivent parfois,
sinon souvent, accepter de reviser leur échelle implicite en se laissant
contraindre par les définitions opérationnelles qui leur sont proposées.
De nouvelles habitudes mentales doivent s'installer, notamment pour
respecter rigoureusement les consignes d'évaluation standardisées.
Une première expérience en milieu scolaire a, en effet, montré que,
sans ce changement personnel, des déformations et des omissions font
perdre une bonne partie des avantages que des échelles rigoureuse-
ment construites peuvent apporter. Pour réduire la difficulté au
maximum, l'utilisation des échelles a été en quelque sorte programmée,
entendons que l'on propose un mode d'emploi pas à pas.
4. La mise au point des échelles proposées ici a demandé un temps
considérable. Incomparablement plus que des enseignants isolés,
si dévoués soient-ils, ne peuvent en investir dans un travail qui ne
constitue pour eux qu'une partie de leur tâche.

Ainsi se confirme, une fois de plus, la nécessité d'instaurer une colla-
boration permanente entre des centres de recherche pédagogique et
les groupes d'enseignants.

Par ailleurs, si des efforts comme celui-ci peuvent continuer, on finira
par disposer d'un vaste choix d'échelles, conçues pour les différentes

branches et pour les principaux types de tâches qu'elles appellent. Entre-temps, de nouvelles habitudes d'évaluation se seront aussi installées chez les éducateurs et chez les élèves (l'auto-évaluation pouvant, en effet, emprunter des chemins similaires à celui-ci). On peut donc espérer que, d'ici quelques années, ce qui est lourd travail aujourd'hui deviendra charge normale.

Nous le savons, les outils proposés vont sembler compliqués aux enseignants qui entreront pour la première fois en contact avec eux. A ces enseignants, nous demandons d'essayer patiemment les échelles, d'en discuter avec leurs collègues, de procéder, par petits groupes, à des expériences d'appréciation comparative de certains travaux. Ils rendront ensuite un grand service à leurs collègues et aux chercheurs en acceptant de leur faire part de leurs critiques, des résultats de leurs essais, de leurs suggestions, soit pour amender les outils proposés, soit pour en construire d'autres.

âches qu'elles appellent.
tion se seront aussi ins-
auto-évaluation pouvant,
celui-ci). On peut donc
ourd travail aujourd'hui

sembler compliqués aux
bis en contact avec eux.
patiemment les échelles,
er, par petits groupes, à
de certains travaux. Ils
ègues et aux chercheurs
es, des résultats de leurs
les outils proposés, soit

I. ... L'ÉVALUATION : QUELQUES CONDITIONS NÉCESSAIRES

✕

✂ Pour qu'elle puisse jouer efficacement son rôle, une évaluation doit d'abord être VALIDE, c'est-à-dire mesurer réellement ce qu'elle prétend mesurer. Pareille affirmation peut sembler banale, sinon naïve. Pourtant, dans la pratique scolaire comme dans la vie courante, combien de fois ne confond-on pas une partie avec le tout (par exemple : évaluer la mémoire en pensant mesurer l'intelligence) ou ne mesure-t-on pas « à côté » (par exemple : poser comme évidemment équivalents le contenu d'un programme scolaire et les exigences de la vie pratique) ? On va d'ailleurs voir que le concept même de *validité* n'est pas absolu, mais relatif : il est indissociable de la finalité poursuivie.

Mais, même si l'on ne fait pas erreur sur l'objet de l'évaluation, celle-ci peut cependant varier de qualité, *l'idéal* (presque jamais atteint) étant d'arriver à *une précision, une OBJECTIVITÉ absolues*.

Enfin, dans un contexte d'éducation, l'évaluation doit présenter des qualités particulières afin de *SERVIR LE MIEUX LES ENSEIGNANTS ET LES ÉLÈVES DANS LEURS TÂCHES RESPECTIVES*.

Des échelles d'évaluation bien construites doivent répondre à toutes ces exigences premières. Nous verrons comment on a essayé d'y satisfaire.

A. — L'ÉVALUATION DOIT ÊTRE VALIDE

Au-delà de nombreuses nuances, on distingue trois types de validité : la *validité prédictive*, la *validité de construct* et la *validité de contenu*.

Imaginons qu'à partir de la combinaison des scores obtenus à un jeu de mots croisés, à un jeu de raisonnement abstrait et à un jeu de puzzle difficile, on puisse prédire les chances de réussite d'un jeune adolescent dans l'enseignement secondaire général. Nous mesurons de toute évidence des comportements cognitifs et des comportements affectifs, mais nous ne savons trop lesquels. Peu nous importe cependant si les prédictions se vérifient dans une mesure satisfaisante. En pareil cas, on dit que l'épreuve possède une bonne *validité prédictive*.

Dans cette démarche, on se soucie plus d'efficacité que de rigueur scientifique, car on ne fait aucun effort particulier pour analyser des processus et découvrir les causes. On sait, toutefois, que mieux on

X

connaît la genèse d'un phénomène, mieux on sait le prédire. Mais, en sciences humaines, nos ignorances fondamentales sont hélas bien plus vastes que notre savoir.

Il en va tout autrement lorsque l'on se soucie réellement de savoir ce que l'on mesure. Rarement, en matière d'éducation, on évalue directement la qualité ou l'objet auxquels on s'intéresse. Mais cela arrive, par exemple quand on mesure la vitesse de lecture, la richesse du vocabulaire, l'exactitude d'une analyse chimique, etc. Le plus souvent, cependant, l'évaluation porte sur des traits psychologiques (attention, intérêts, attitudes,...) ou des aptitudes (intelligence verbale, créativité,...). Alors que, pour construire un concept, nous dégageons progressivement les qualités communes à un plus ou moins grand nombre d'objets que nous avons directement observés, il n'en va pas de même pour les traits psychologiques ou les aptitudes.

Nous ne pouvons placer l'intelligence sous le microscope, ni dire de quelle couleur est la créativité. L'intelligence, comme la créativité, nous ne les voyons jamais, nous ne les connaissons qu'à travers certains comportements qui, en raison de caractéristiques communes, semblent dépendre d'un même « poste de commande », d'un même « ordinateur ». Mais il ne s'agit là que d'entités supposées, de concepts construits non plus à partir de l'observation de ces « postes », de ces « ordinateurs », mais des manifestations qu'on leur attribue. Des concepts ainsi « reconstruits » sont appelés *constructs*.

Selon Thorndike et Hagen ⁽¹⁾, la validité de construct d'un instrument de mesure est établie quand il permet d'obtenir des informations significatives sur les individus mesurés. Il va de soi que la validité de construct sera d'autant plus élevée qu'elle tend à l'exploration exhaustive de l'entité considérée : idéalement, un test d'intelligence verbale doit mesurer *tous* les aspects de ce type d'intelligence.

Mais, avant de se soucier d'évaluation exhaustive, l'enseignant doit se préoccuper d'éducation complète. Entendons qu'idéalement au moins, il est désirable de cultiver toutes les aptitudes de l'élève. Cependant, un double obstacle surgit : d'abord, les aptitudes d'un individu normal

(1) R. L. THORNDIKE et E. HAGEN, *Measurement and evaluation in psychology and education*, New York, Wiley, 1969, 3^e éd.

sont tellement nombreux
toutes avec le même
on décide de s'attacher
lenteur. L'école d'aujourd'hui
des aptitudes des élèves
que d'investir l'essentiel

La validité de contenu
éducatives et, pendant le
en termes de matières d'
enseignement ou une év
de validité de contenu,
ou du plan d'études off
d'arithmétique, destiné à
programme prévoit, on
mentionnées dans les do
ce que signifie être bie
(question qui relève de l

Mais cette façon de c
Les matières ne sont qu
fication que si des com
œuvre. La validité de c
comportements à appr
taxonomies comme celle
tion : elles nous invitent
et constituent une mise
niveau de la simple mém

A propos de validité
celui de la cohérence en
il arrive que l'on souhai
soucier des sources de s
situation d'enseigne
l'action de l'éducateur
des sanctions à un élève
choses auxquelles on ne
en décide autrement de f
eux doit exister entre l

sont tellement nombreuses qu'il est bien difficile de les développer toutes avec le même soin; ensuite, beaucoup d'aptitudes auxquelles on décide de s'attacher plus spécialement, ne se développent qu'avec lenteur. L'école d'aujourd'hui agit donc sagement en se souciant plus des aptitudes des élèves et en concevant son action à long terme, plutôt que d'investir l'essentiel de son énergie dans des savoirs momentanés.

La *validité de contenu* représente une troisième approche des réalités éducatives et, pendant longtemps, elle a été trop exclusivement conçue en termes de matières d'enseignement. Par exemple, on estimait qu'un enseignement ou une évaluation satisfaisaient aux exigences en matière de validité de contenu, si tous les points importants du programme ou du plan d'études officiels étaient convertis. Pour construire un test d'arithmétique, destiné à contrôler si les élèves ont bien appris ce que le programme prévoit, on va donc procéder à un inventaire des matières mentionnées dans les documents officiels, et non se demander vraiment ce que signifie être bien éduqué dans le domaine de l'arithmétique (question qui relève de la validité de construct).

Mais cette façon de concevoir la validité de contenu est trop étroite. Les *matières* ne sont que matériaux inertes; elles ne prennent de signification que si des comportements se produisent pour les mettre en œuvre. La validité de contenu implique donc aussi un inventaire des comportements à apprendre ou à perfectionner. C'est en quoi des taxonomies comme celle de Bloom rendent tant de services à l'éducation : elles nous invitent à un dépassement cognitif ou affectif continu et constituent une mise en garde permanente contre la stagnation au niveau de la simple mémorisation des faits et contre la passivité.

A propos de validité de contenu, un autre problème se pose encore : celui de la cohérence entre l'enseignement et l'évaluation. Assurément, il arrive que l'on souhaite évaluer ce que sait ou peut un élève, sans se soucier des sources de son savoir ou de son pouvoir. Mais, dans une situation d'enseignement, il est bon de reconnaître dans quelle mesure l'action de l'éducateur a été efficace... et il est inadmissible d'infliger des sanctions à un élève qui se révélerait incapable d'accomplir des choses auxquelles on ne l'a pas préparé. Bref, et à moins que l'on en décide autrement de façon tout à fait explicite, un parallélisme rigoureux doit exister entre le contenu de l'enseignement et le contenu de

X

l'évaluation. Il est, par exemple, inadmissible de gaver la mémoire pendant l'enseignement, puis de poser des questions dites d'intelligence lors de l'évaluation, c'est-à-dire d'inviter alors les élèves à travailler à un niveau cognitif auquel ils n'ont pas appris à se mouvoir. Or plus un comportement est complexe, plus il appelle de soins éducatifs. Prétendre que les apprentissages les plus nobles relèvent de processus tellement subtils qu'on ne peut ni les cultiver systématiquement ni les évaluer objectivement, c'est tout simplement éluder la difficulté... et ses responsabilités d'éducateur.

Mais l'exigence de parallélisme entre l'enseignement et l'évaluation en entraîne une autre : *la participation indispensable des enseignants à la construction des instruments d'évaluation*. Car eux seuls savent réellement ce qui se passe dans leur classe. Nous traversons une période de transition où la rigueur et la complexité accrues de l'évaluation rend nécessaire l'intervention de personnes spécialement formées, mais il faut que, le plus rapidement possible, deux objectifs soient atteints : que les enseignants acquièrent les connaissances, la maîtrise des techniques nouvelles et qu'une gamme complète d'instruments ou de modèles d'instruments soit mise à leur disposition.

Il semble cependant probable qu'une *collaboration permanente* va s'établir *entre enseignants et chercheurs*. *La construction des échelles d'évaluation descriptives, décrites dans le présent ouvrage, a permis de démontrer, une fois de plus, combien pareille coopération est fructueuse pour tous.*

Faut-il opter nécessairement entre la validité prédictive, la validité de construct et la validité de contenu ? Certes non. *L'idéal est, au contraire, qu'un instrument soit valide à tous égards*. On peut, par exemple, espérer que les échelles d'évaluation qui nous occupent explorent bien l'aptitude à s'exprimer par écrit (validité de construct), que cette exploration répond au souci de responsables de l'enseignement rénové et s'étend à tous les comportements importants (validité de contenu) et, enfin, qu'elle permet de prédire dans quelle mesure les élèves ainsi évalués pourront s'exprimer à l'avenir (validité prédictive).

B. — L'ÉVALUATION

Une évaluation est répétitive, fidèle d'un objet, presque jamais atteinte de la complexité des objets évalués, presque inévitablement soumise à un subterfuge : la mesure, on se contente de la constance des résultats (lité). On est en droit de dire, par habitude, il arrive souvent de même façon et que leur résultat est unanime que d'une objection.

D'abord définir, de façon précise, la condition de la validité pour les raisons des discordances que quelques-unes : le problème voudra bien se résoudre.

En l'absence de tout jugement, se basant sur les performances de leurs élèves, se donner une même performance et insuffisante dans un autre.

Même si un accord peut être obtenu dans l'évaluation, il est de juge à juge. Les écarts de l'élève est complexe.

En outre, la note peut être étrangère à la performance due à une mauvaise

Certains procédés peuvent être faibles : correction à une même question,

B. — L'ÉVALUATION DOIT, AUTANT QUE POSSIBLE, ÊTRE OBJECTIVE

Une évaluation est réputée objective si elle fournit une mesure impartiale, fidèle d'un objet. En sciences humaines, pareille qualité n'est presque jamais atteinte de façon parfaite, non seulement en raison de la complexité des objets évalués, mais aussi de l'interaction qui se produit presque inévitablement entre évaluateur et évalué. Aussi recourt-on souvent à un subterfuge : faute de pouvoir être certain de la qualité de la mesure, on se contente de l'accord entre mesureurs (concordance) ou de la constance des mesures effectuées par un même individu (fidélité). On est en droit de penser que par conditionnement commun, par habitude, il arrive souvent que les évaluateurs se trompent de la même façon et que leur jugement témoigne plutôt d'une subjectivité unanime que d'une objectivité de caractère absolu. Mais que faire ?

D'abord définir, de façon aussi précise que possible, l'objet de l'évaluation, condition de la validité de construct. Ensuite, s'efforcer de découvrir les raisons des discordances entre évaluateurs. Nous n'en rappelons que quelques-unes : le lecteur plus spécialement intéressé par ce problème voudra bien se reporter aux principaux traités de docimologie.

En l'absence de toute concertation sur les objectifs poursuivis, des juges, se basant sur leur expérience personnelle et sur les possibilités de leurs élèves, se donnent des critères différents et parfois peu stables. Une même performance peut donc être estimée excellente dans un groupe et insuffisante dans un autre.

Même si un accord peut être atteint sur les principaux critères à retenir dans l'évaluation, l'importance qui leur est accordée varie souvent de juge à juge. Les écarts sont d'autant plus accusés que la performance de l'élève est complexe.

En outre, la note peut être contaminée par l'intervention de facteurs étrangers à la performance évaluée : charme exercé par l'élève, irritation due à une mauvaise écriture, préjugés de l'évaluateur, etc.

Certains procédés permettent de pallier, au moins en partie, pareilles faiblesses : correction anonyme, correction en série de toutes les réponses à une même question, variation de l'ordre de correction des copies, etc.

Mais, au-delà de ces antidotes, une démarche plus fondamentale s'impose. Elle consiste essentiellement dans la définition commune et opérationnelle des objectifs de la tâche proposée, y compris la définition des critères (qualitatifs et quantitatifs) de l'évaluation. Les instruments présentés dans cet ouvrage sont le fruit d'un tel effort de précision et de concertation.

INFORMER L'ÉLÈVE

Dans une situation d'apprentissage, l'évaluation sert essentiellement à informer l'élève de la qualité de ses performances, de son progrès vers l'objectif poursuivi. Depuis longtemps déjà, on connaît le rôle important de cette information. Grâce surtout aux travaux théoriques et expérimentaux de B. F. Skinner, on sait aujourd'hui que ce rôle est déterminant ⁽¹⁾.

Est-il encore besoin de souligner que l'évaluation est d'autant plus aisée que l'objectif à atteindre est défini avec précision. Et cet objectif, l'élève doit naturellement le connaître, le comprendre, et l'avoir fait sien.

INFORMER L'ENSEIGNANT

L'évaluation est aussi nécessaire à l'enseignant. Sans elle, il ne peut savoir dans quelle mesure son action a été efficace, ni comment il doit la modifier à l'avenir pour mieux aider l'élève.

Pour remplir efficacement ces fonctions, l'évaluation doit satisfaire à un certain nombre d'exigences qui s'ajoutent à celles que nous avons déjà rencontrées.

C. — L'ÉVALUATION DOIT ÊTRE FORMATIVE

Cette exigence vient déjà d'être formulée de façon générale, mais le point est suffisamment important pour que l'on s'y arrête un instant.

Trop souvent, dans le passé, l'évaluation s'est réduite à un classement transformant la vie scolaire en une sorte de concours perpétuel. Deux soucis dominaient cette façon de faire : découvrir les meilleurs (et, si

(1) Rappelons que B. F. SKINNER définit le *renforcement* comme tout événement qui, survenant à la suite d'une réponse, en augmente la probabilité d'émission.

B. F. SKINNER, *L'analyse expérimentale du comportement*, Bruxelles, Dessart, 1971.

possible, l'élite) et stimuler, la concurrence a presque certainement été oubliée. Outre, l'appréciation a été oubliée. Quelles informations venons-nous moins bien qu'un autre ? L'objectif à atteindre, peut-on ne peut plus avancer ? La seconde conception de l'évaluation, la formation première conception sert, et les autres. Assurément, ce n'est pas un dépouillage d'intérêt, mais une bonne psychologie de l'évaluation éventuellement le dernier.

Dans une situation de difficulté à surmonter, le processus mis en œuvre par l'élève et son

Lorsque nous nous engageons dans un travail, nous attendons des résultats relatifs à la tâche et nous nous aidons à progresser. Par conséquent, notre conseiller mettra l'accent sur le Bref, l'évaluation doit être basée sur le travail et non sur la personne. Le travail que l'on est soi-même

Les études documentaires complexes ont maintes fois abouti en suivant un processus de classement d'une note globale. Les choses ne sont peut-être pas si lourdes de la démarche par une note beaucoup plus une fois pour toutes,

plus fondamentale s'impose.
commune et opérationnelle
la définition des critères
Les instruments présentés
effort de précision et de

ation sert essentiellement
performances, de son progrès
déjà, on connaît le rôle
t aux travaux théoriques
aujourd'hui que ce rôle est

évaluation est d'autant plus
précision. Et cet objectif,
prendre, et l'avoir fait sien.

ant. Sans elle, il ne peut
ace, ni comment il doit la

évaluation doit satisfaire
t à celles que nous avons

E FORMATIVE

façon générale, mais le
on s'y arrête un instant.

t réduite à un classement
concours perpétuel. Deux
ouvrir les meilleurs (et, si

ment comme tout événement
probabilité d'émission.
rtement, Bruxelles, Dessart,

possible, l'élite) et stimuler chacun par la compétition. Malheureusement, la concurrence a surtout encouragé les premiers (qui l'auraient presque certainement été de toute façon) et découragé les autres. En outre, l'appréciation ainsi formulée ne servait guère l'apprentissage. Quelles informations veut-on donner à l'élève? S'il travaille mieux ou moins bien qu'un autre, ou bien à quelle distance il se trouve encore de l'objectif à atteindre, pourquoi il n'est pas plus loin, voire pourquoi il ne peut plus avancer? Pour aider vraiment chacun à progresser, cette seconde conception de l'évaluation s'impose. Elle sert directement l'éducation, la formation (d'où le nom d'évaluation formative). La première conception sert, elle, à informer, aide à se situer par rapport aux autres. Assurément, ce type d'information est loin d'être toujours dépourvu d'intérêt, mais on conviendra facilement qu'il n'est pas de très bonne psychologie de répéter, jour après jour, à un élève qu'il est le x^e , éventuellement le dernier de sa classe.

Dans une situation d'apprentissage, il importe que la saine notion de difficulté à surmonter se substitue à la notion d'échec (si souvent dramatisé par l'élève et ses parents).

Lorsque nous nous entourons de conseils d'autrui pour mieux effectuer un travail, nous attendons des suggestions, des indications, des critiques relatives à la tâche et nous les acceptons d'autant mieux qu'elles nous aident à progresser. Par contre, il ne nous plaît guère qu'à cette occasion, notre conseiller mette en cause notre personnalité, notre intelligence, ... Bref, l'évaluation doit porter sur l'objet, sur l'apprentissage à réaliser, et non sur la personne. Ce n'est pas parce que l'on a fait du mauvais travail que l'on est soi-même mauvais.

Les études docimologiques sur la notation des performances complexes ont maintes fois démontré que la note à laquelle un correcteur aboutit en suivant une démarche analytique ne diffère pas nécessairement d'une note globale, accordée en un temps parfois bref. Les choses ne sont peut-être pas aussi simples, mais il est indéniable que la lourdeur de la démarche analytique est loin d'être toujours récompensée par une note beaucoup plus valide. Est-ce une raison pour renoncer, une fois pour toutes, à la méthode analytique?

L'ÉVALUATION
DOIT PORTER SUR
LA PERFORMANCE
ET NON SUR LA
PERSONNE

L'ÉVALUATION
FORMATIVE DOIT
ÊTRE ANALYTIQUE

X

Pas du tout. Car les deux démarches ne servent pas le même propos. A quoi sert-il de dire à un élève que son travail est « bon » ou « faible » sans lui montrer pourquoi il en est ainsi? En particulier, comment celui qui apprend pourra-t-il surmonter les difficultés qu'il éprouve si elles ne sont pas précisées, localisées (en vue d'être systématiquement attaquées)?

Bref, l'évaluation formative est essentiellement analytique, diagnostique. Le critère du choix entre notation analytique et notation globale ne peut donc être la rapidité, la convenance du correcteur, mais bien l'objectif éducatif poursuivi.

**LES OBJECTIFS
DOIVENT ÊTRE
OPÉRATIONALISÉS**

Un pédagogue anglais déclarait récemment avec malice que si l'on sait où l'on va, on a plus de chances d'y arriver que si on l'ignore... Les objectifs de l'éducation doivent donc être précisés. En outre, comme on ne peut tous les poursuivre en même temps, il importe de les ordonner. Dans certains cas, et une fois l'objectif général précisé (par exemple, apprendre à résoudre certains problèmes de la vie courante par le calcul algébrique), l'ordonnance des objectifs particuliers est en bonne partie dictée par la structure de la matière (il faut connaître la règle des signes pour pouvoir recourir au calcul algébrique). Dans d'autres cas, l'ordre ne s'impose pas avec une telle force.

Plus on s'éloigne des apprentissages de base, plus on s'élève dans la vie intellectuelle et affective, plus les cheminements de l'apprentissage se diversifient.

L'une des tâches essentielles de l'enseignant est de découvrir les nœuds, les « plaques tournantes », les points critiques et de rechercher les stratégies qui optimiseront l'acquisition de ces savoirs ou de ces pouvoirs fondamentaux pour la majorité des élèves, idéalement pour tous.

Répetons-le, dans cette perspective, une évaluation qui prend pour critère la performance d'autrui n'a guère de sens. Ce qui importe, c'est d'informer l'élève de son progrès vers l'objectif poursuivi. En pareil cas, il faut y insister, l'évaluation peut pleinement jouer son rôle de régulation, de renforcement, car elle est partie intégrante du processus d'apprentissage et trouve l'objectivité de son critère dans la plus ou moins bonne réussite de l'entreprise.

On imagine combien hélas encore fréquent, ces classements comparatifs de mauvaise qualité, mais à celle des autres. L'enseignant sélective, d'émulation acceptent difficilement les élèves (?). Pourtant, un celui qui conduit tous les

Continue, analytique trois qualités indispensables guider l'élève dans son les informations dont il de ces qualités, l'évaluation

Le triple souci de dans les directives officielles qu'elle s'est initialement rénové répond-elle à ce a eu pour objectifs de d'étudier les réactions d'application des nouvelles

(?) On rencontre encore la note minimale pour réussir. L'évaluateur peut ainsi pe

ent pas le même propos.
l est « bon » ou « faible ».
En particulier, comment
difficultés qu'il éprouve si
d'être systématiquement

lement analytique, dia-
alytique et notation glo-
nce du correcteur, mais

avec malice que si l'on
ver que si on l'ignore...
écisés. En outre, comme
il importe de les ordon-
ral précisé (par exemple,
ie courante par le calcul
iers est en bonne partie
naître la règle des signes
ans d'autres cas, l'ordre

plus on s'élève dans la
ments de l'apprentissage

nt est de découvrir les
critiques et de rechercher
ces savoirs ou de ces
élèves, idéalement pour

luation qui prend pour
s. Ce qui importe, c'est
tif poursuivi. En pareil
ment jouer son rôle de
intégrante du processus
re dans la plus ou moins

On imagine combien pareille façon de procéder diffère de l'usage hélas encore fréquent, où une performance est dépréciée par le jeu des classements comparatifs, non pas nécessairement parce qu'elle est de mauvaise qualité, mais parce qu'elle est inférieure, parfois de bien peu, à celle des autres. L'enseignement est tellement imprégné de préoccupation sélective, d'émulation artificielle, que beaucoup de professeurs acceptent difficilement l'idée d'attribuer une note très bonne à tous les élèves (1). Pourtant, un enseignement fondamental réussi n'est-il pas celui qui conduit tous les élèves à bon port ?

Continue, analytique, centrée sur l'apprentissage à réaliser sont trois qualités indispensables pour que l'évaluation serve réellement à guider l'élève dans son progrès quotidien, et à donner à l'enseignant les informations dont il a besoin pour mieux adapter son action. Faute de ces qualités, l'évaluation ne remplit pas son rôle *formatif*.

* * *

Le triple souci de validité, d'objectivité et de précision apparaît dans les directives officielles. L'échelle d'évaluation à cinq degrés telle qu'elle s'est initialement répandue dans l'enseignement secondaire rénové répond-elle à ces exigences cruciales ? Une enquête préliminaire a eu pour objectifs de répondre à cette question et, plus généralement, d'étudier les réactions des professeurs après une première période d'application des nouveaux modes d'évaluation.

(1) On rencontre encore des cas, notamment dans l'enseignement supérieur, où la note minimale pour réussir un examen n'est fixée qu'après correction des travaux. L'évaluateur peut ainsi perpétuer la distribution traditionnelle des notes.

II. ENQUÊTE SUR L'ÉVALUATION

Le système d'évaluation à a
 rénové a été présenté et justif
 La rénovation se caractérise c
 tionnels au profit d'un contr
 grante de l'apprentissage. L'c
 d'adapter constamment le trav
 d'aboutir à une meilleure indi
 Il s'agit donc bien d'une évalu

La seconde caractéristique
 additions de points et des cl
 points (1970) de cristalliser une
 tion de tels points déforme la r
 de l'élève pèse continuellement
 contre des objectifs des cycles
 permettre de suivre l'évolution
 leur guidance dans le déroul
 les erreurs d'aiguillage et les pe
 La subjectivité de la notation c
 le manque de concordance ent
 et le danger de stéréotypie de
 sur ses élèves. En remplacem
 une notation qualitative, sele
 sant, Faible, Insuffisant) ⁽²⁾,
 tout exercice ou travail d'élé

(¹) Voir notamment :

— Circulaire du 14 octobre 1969
 — Evaluation continue et contr
renové. Organisation générale du
 1970.

— Modalités d'appréciation da
point sur l'enseignement secondaire
 Etudes du Ministère de l'Education

(²) Evaluation continue et contr

(³) Lorsque ces cinq niveaux son
 d'échelle littérale.

A. — CONTEXTE DE L'ENQUÊTE

Le système d'évaluation à appliquer dans l'enseignement secondaire rénové a été présenté et justifié dans plusieurs documents officiels (1). La rénovation se caractérise d'abord par l'abandon des examens traditionnels au profit d'un contrôle des connaissances faisant partie intégrante de l'apprentissage. L'objectif est de « permettre au professeur d'adapter constamment le travail de la classe à l'état réel de celle-ci et d'aboutir à une meilleure individualisation de l'enseignement » (1969). Il s'agit donc bien d'une évaluation formative.

La seconde caractéristique est l'abandon de la note chiffrée, des additions de points et des classements. On reproche au système des points (1970) de cristalliser une situation momentanée, tandis que l'addition de tels points déforme la réalité et entraîne des décisions où le passé de l'élève pèse continuellement sur son avenir. Cette pratique va à l'encontre des objectifs des cycles d'observation et d'orientation qui doivent permettre de suivre l'évolution des élèves et de leur « assurer une meilleure guidance dans le déroulement de leurs études afin de leur éviter les erreurs d'aiguillage et les pertes de temps qui peuvent en résulter » (2). La subjectivité de la notation chiffrée est dénoncée et, plus spécialement, le manque de concordance entre les évaluations de différents professeurs et le danger de stéréotypie des jugements portés par un même maître sur ses élèves. En remplacement du système traditionnel, on propose une notation qualitative, selon cinq niveaux (Très bien, Bien, Satisfaisant, Faible, Insuffisant) (3). Ce type de notation permet d'évaluer tout exercice ou travail d'élève dans le courant de l'année scolaire.

(1) Voir notamment :

- Circulaire du 14 octobre 1969. Modalités d'appréciation.
- Evaluation continue et contrôle des connaissances, in « *Enseignement secondaire rénové. Organisation générale du 1^{er} degré* », Ministère de l'Éducation Nationale, 1970.
- Modalités d'appréciation dans l'enseignement secondaire rénové, in *Faire le point sur l'enseignement secondaire rénové. Organisation générale*, Organisation des Etudes du Ministère de l'Éducation Nationale, 1974.

(2) Evaluation continue et contrôle des connaissances, *op. cit.*

(3) Lorsque ces cinq niveaux sont symbolisés par les lettres A, B, C, D, E, on parle d'échelle littérale.

Les appréciations portées quatre fois par an dans le bulletin, concernent trois opérations mentales essentielles : les connaissances, la compréhension et l'utilisation de l'acquis. Une rubrique supplémentaire « Observation et conseils éducatifs » est réservée à l'évaluation des autres processus comme la capacité d'analyse, de synthèse, d'invention, le sens critique, ainsi qu'à d'autres notations concernant le comportement psychosocial.

La notation qualitative est recoupée par des contrôles à base quantitative suivant les principes de la statistique. On a, par exemple, préconisé que, deux ou trois fois par an, les professeurs d'une même discipline proposent à toutes les classes du même niveau un exercice de contrôle noté sur cent. La note moyenne peut alors servir d'axe de distribution de l'ensemble des élèves du niveau envisagé⁽¹⁾. Cette démarche vise à ajouter une référence normative à l'évaluation continue, de nature formative.

« Dans un système où sont abolis les points et les classements, le conseil de classe prend une importance majeure. Il doit viser plus que jamais à dégager une image évolutive de chaque enfant et à comprendre les raisons des faiblesses ou des difficultés constatées. »⁽²⁾ Cette image se traduit dans deux documents, l'un destiné aux élèves et à leurs parents (le bulletin), l'autre au corps professoral (le dossier scolaire).

Dès la parution de ces instructions officielles, de nombreux enseignants, encouragés par les inspecteurs et les chefs d'établissement, se sont résolument mis au travail pour essayer honnêtement un système dont ils approuvaient la philosophie. A bien des égards, le nouveau système de notation était loin de l'ancien : dans sa présentation, il était souvent fait référence à des notions de statistique descriptive peu familières aux enseignants, et bien d'autres difficultés encore les attendaient. Mais n'anticipons pas : l'important est de souligner le caractère favorable de l'accueil initial et la bonne volonté qui a présidé aux premiers essais.

⁽¹⁾ Evaluation continue et contrôle des connaissances, *op. cit.*

⁽²⁾ *Idem.*

B. — OBJECTIFS

Deux ans après l'introduction du nouveau système, il a semblé utile de recueillir des impressions.

- D'une part, des informateurs ont été interrogés : les professeurs utilisaient-ils le nouveau système ? contrôlaient-ils les apprentissages ? évaluaient-ils leurs évaluations ? les résultats de leurs élèves ?
- D'autre part, des informateurs ont été interrogés : les professeurs des autres disciplines ont-ils adopté le nouveau système ?

Depuis l'enquête, plusieurs établissements ont resté cependant d'une actualité particulière. Ce rappel paraît d'autant plus intéressant que de nombreux enseignants belges ont dû adopter aussi le nouveau système.

C. — POPULATION

Un questionnaire a été adressé aux professeurs de français et de mathématique. Ce choix a été motivé par :

- 1° L'importance des deux disciplines ;
- 2° L'impossibilité matérielle de faire un questionnaire plus large ;
- 3° La nécessité de disposer de données de référence dans les deux disciplines.

Sur les 89 établissements interrogés, 391 professeurs ont répondu. Parmi eux, 194 professeurs de français et 197 professeurs de mathématique.

B. — OBJECTIFS DE L'ENQUÊTE

Deux ans après l'introduction du nouveau système d'évaluation, il a semblé utile de recueillir deux types d'informations :

- D'une part, des informations objectives : Quel type d'échelle les professeurs utilisaient-ils ? A quelle fréquence et sous quelle forme contrôlaient-ils les apprentissages de leurs élèves ? Comment synthétisaient-ils leurs évaluations en fin de période ? Comment situaient-ils les résultats de leurs élèves ?...
- D'autre part, des informations subjectives : essentiellement, l'avis des professeurs sur la difficulté, les avantages et les inconvénients du nouveau système.

Depuis l'enquête, plusieurs années se sont écoulées. Ses résultats restent cependant d'une actualité suffisante pour que nous les rappelions. Ce rappel paraît d'autant plus s'imposer qu'à partir de septembre 1976, de nombreux enseignants belges, qui n'avaient pas encore été directement atteints par les nouvelles instructions relatives à l'évaluation, ont dû adopter aussi le système rénové.

C. — POPULATION ET DISCIPLINES CONCERNÉES

Un questionnaire a été adressé aux professeurs de langue maternelle et de mathématique. Ce choix s'explique par les raisons suivantes :

- 1° L'importance des deux branches considérées.
- 2° L'impossibilité matérielle d'étendre l'enquête à tous les professeurs.
- 3° La nécessité de disposer, pour une phase ultérieure de la recherche, de données de référence qui n'existent actuellement que pour ces deux disciplines.

Sur les 89 établissements d'enseignement secondaire qui avaient entrepris la rénovation au moment de l'enquête, 80 ont participé. Au total, 391 professeurs ont renvoyé leurs réponses, soit 197 professeurs de langue maternelle et 194 professeurs de mathématique.

D. — QUESTIONNAIRE

Sept rubriques ont été retenues :

- Modes de contrôles.
- Fréquence des contrôles.
- Niveaux taxonomiques des contrôles.
- Synthèse des évaluations en fin de période.
- Coordination des contrôles.
- Références utilisées.
- Difficultés liées au nouveau système d'évaluation.

Les questions ont été rédigées sous leur forme définitive après essais auprès d'un nombre limité de professeurs.

E. — DÉPOUILLEMENT

Le dépouillement s'est effectué en trois phases :

- 1^o Codage des réponses aux questions fermées.
- 2^o Analyse de contenu visant à dégager les thèmes les plus fréquents des réponses aux questions ouvertes. Calcul de la fréquence d'apparition de ces thèmes.
- 3^o Traduction graphique de ces informations.

F. — RÉSULTATS

1. — MODES DE CONTRÔLES

1.a.1. UTILISEZ-VOUS L'ÉCHELLE À CINQ DEGRÉS : TB-B-S-F-I ?

Tous les professeurs utilisent l'échelle à cinq degrés imposée par le Ministère.

1.a.2. CERTAINS PROFESSEURS C NOTATION. SI C'EST VOTRE

35 professeurs de mathématique nuancent l'échelle à cinq

- soit en ajoutant deux degrés (Nul) ; on aboutit ainsi à un
- soit en fractionnant un ou deux degrés. Exemple : S⁺, S⁼, S⁻
- soit encore :
 - en situant la note entre deux
 - en utilisant des signes de couleur :
 - vert = travail valable globalement
 - bleu = travail incomplet
 - orange = travail très incomplet

L'échelle imposée ne satisfait pas généralement parce qu'elle manque de précision (« Le nouveau système d'évaluation est-il plus précis ? ») que les commentaires

1.b.1. ÉVALUEZ-VOUS LE TRAVAIL QUE VOUS TRANSPOSEZ

62 % des professeurs de langue maternelle évaluent le travail ensuite dans une nouvelle échelle

Ces pourcentages élevés montrent un simplisme d'y voir la manifestation d'un changement. L'habitude joue toutefois à penser que le système est un instrument de pensée plus précis et d'ordination.

Le cas des exercices qui

1.a.2. CERTAINS PROFESSEURS ONT ÉLABORÉ LEUR PROPRE SYSTÈME DE NOTATION. SI C'EST VOTRE CAS, DÉCRIVEZ-LE.

35 professeurs de mathématique et 29 professeurs de langue maternelle nuancent l'échelle à cinq degrés :

- soit en ajoutant deux degrés extrêmes (exemple : Parfait ou excellent, Nul) ; on aboutit ainsi à une échelle à sept degrés;
- soit en fractionnant un ou plusieurs degrés (Création de sous-catégories. Exemple : S⁺, S⁼, S⁻);
- soit encore :
 - en situant la note entre deux degrés dans le bulletin,
 - en utilisant des signes de couleur à chaque question :
 - vert = travail valable globalement
 - bleu = travail incomplet ou partiellement valable ou moyen
 - orange = travail très incomplet ou non satisfaisant.

L'échelle imposée ne satisfait donc pas tous les professeurs, principalement parce qu'elle manque de nuances. On verra à la question 6.a.1 (« Le nouveau système d'évaluation vous paraît moins précis que l'ancien ? ») que les commentaires vont dans ce sens.

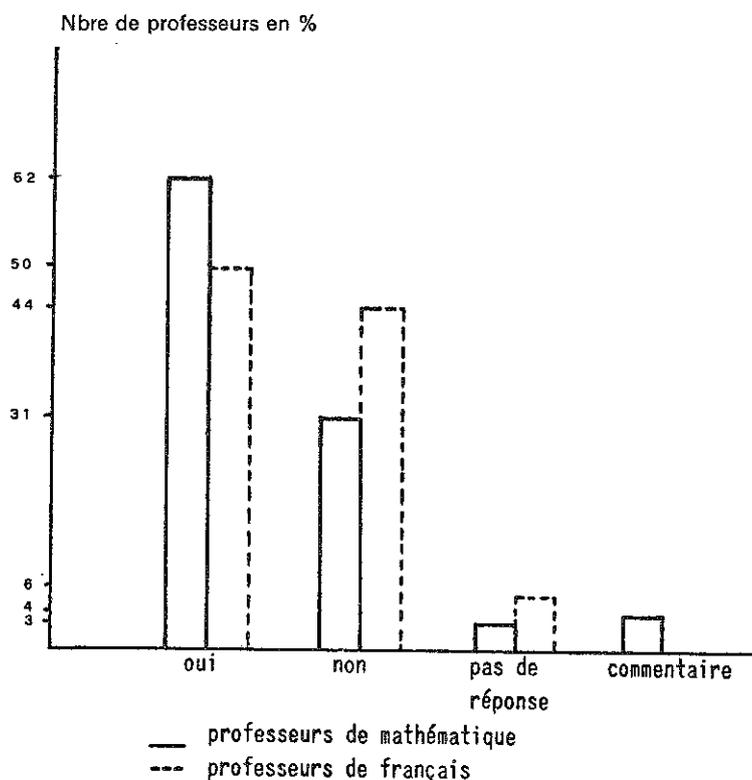
1.b.1. ÉVALUEZ-VOUS LE TRAVAIL D'UN ÉLÈVE PAR UNE NOTE CHIFFRÉE QUE VOUS TRANSPOSEZ ENSUITE DANS LA NOUVELLE ÉCHELLE ?

62 % des professeurs de mathématique et 50 % des professeurs de langue maternelle évaluent par une note chiffrée qu'ils transposent ensuite dans une nouvelle échelle (voir graphique page suivante).

Ces pourcentages élevés méritent l'attention. Il serait, croyons-nous, simpliste d'y voir la manifestation d'une résistance condamnable au changement. L'habitude joue certes un certain rôle. Nous inclinons toutefois à penser que le symbole chiffré constitue chez beaucoup un instrument de pensée plus significatif que la lettre, au cours du processus d'ordination.

Le cas des exercices qui se corrigent selon un décompte rigoureux

(établissement d'un score) doit être traité à part. Pareil exercice peut, en effet, être simultanément affecté d'un score et d'une note d'évaluation.

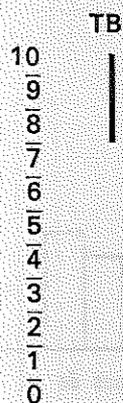


Mais, à côté de la démarche finale qui conduit d'un nombre à une appréciation globale, il importe d'examiner les critères sur lesquels l'évaluation première se fonde. Si ces critères ne sont pas définis de telle sorte que chacun puisse les adopter et s'y référer de la même façon, on risque de retrouver toute la gamme des relativations locales et circonstanciées que la docimologie a, maintes fois, mises en lumière.

A cet égard, les limites numériques des divers degrés, signalées par certains professeurs, apportent une indication claire.

- La limite inférieure de l'échelon
- La limite supérieure de l'échelon inférieure de l'échelon
- La limite supérieure de l'échelon inférieure de l'échelon
- La limite supérieure de l'échelon inférieure de l'échelon
- La limite supérieure de l'échelon inférieure de l'échelon

Le graphique ci-dessous notes :



On constate donc que, selon la notation, 8/10 est traduit par TB, B ou 8; 6/10 est traduit par B, S ou F; 4/10 est traduit par S, F ou L.

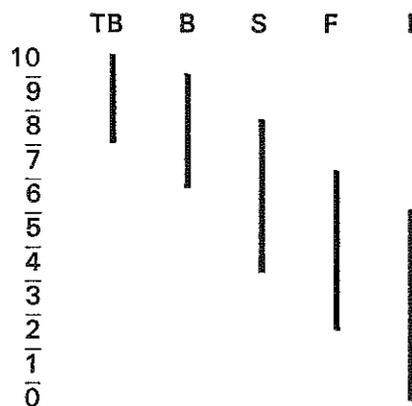
Dans l'état actuel des configurations des notes attribuées, il est difficile de se prononcer sur l'homogénéité d'un même degré de l'échelle.

La nécessité d'une échelle de notation des notes (modération)

té à part. Pareil exercice peut,
core et d'une note d'évaluation.

La limite inférieure de l'échelon	TB	va de 7,5 à 9
La limite supérieure de l'échelon	B	va de 7,5 à 9
inférieure de l'échelon	B	va de 6 à 7,5
La limite supérieure de l'échelon	S	va de 6 à 8
inférieure de l'échelon	S	va de 4 à 7
La limite supérieure de l'échelon	F	va de 4 à 6,5
inférieure de l'échelon	F	va de 2 à 5
La limite supérieure de l'échelon	I	va de 2 à 5
inférieure de l'échelon	I	va de 0 à 2

Le graphique ci-dessous montre clairement le chevauchement des notes :



On constate donc que, selon les professeurs,

8/10 est traduit par TB, B ou S

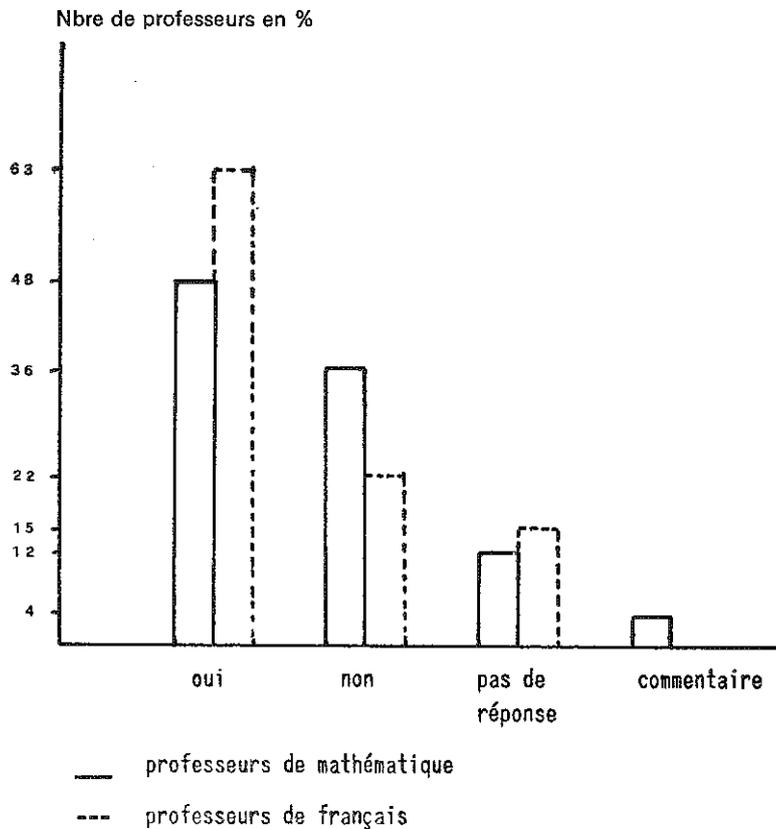
6/10 est traduit par B, S ou F

4/10 est traduit par S, F ou I.

Dans l'état actuel des consignes fournies aux professeurs, la signification des notes attribuées manque donc de clarté. Il est impossible de se prononcer sur l'homogénéité des comportements rapportés à un même degré de l'échelle.

La nécessité d'une échelle descriptive et d'une procédure d'harmonisation des notes (modération) semble donc évidente.

1.b.2. ÉVALUEZ-VOUS LE TRAVAIL D'UN ÉLÈVE PAR UNE APPRÉCIATION GLOBALE DE LA PERFORMANCE ?

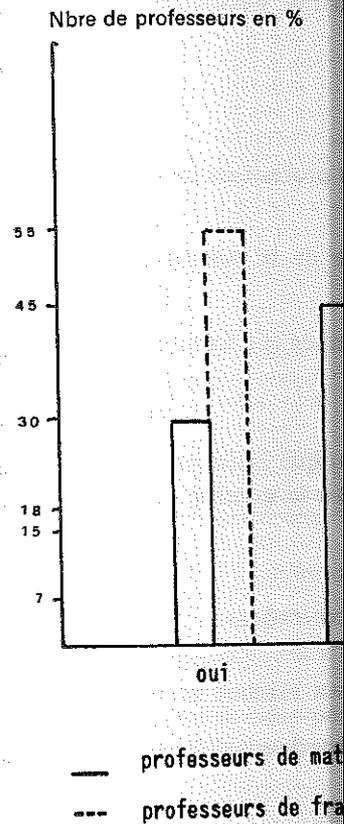


48 % des professeurs de mathématique et 63 % des professeurs de langue maternelle évaluent le travail des élèves par une appréciation globale de la performance. La répartition des pourcentages est inversée par rapport à la question précédente (1.b.1).

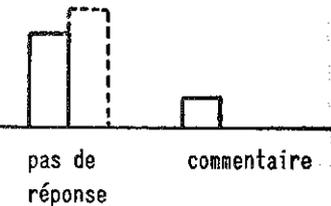
Il semblerait qu'un grand nombre de professeurs utilisent simultanément la note chiffrée et l'appréciation globale. La comparaison des deux graphiques (1.b.1 et 1.b.2) laisse cependant supposer que ces deux modes d'évaluation peuvent s'exclure dans l'esprit de certains enseignants.

- . Une partie des professeurs la note chiffrée.
- . Une partie des professeurs de l'appréciation globale.

1.b.3. ÉVALUEZ-VOUS LE TRAVAIL D'UN ÉLÈVE PAR UNE NOTE CHIFFRÉE SAUF UN CERTAIN NOMBRE D'ÉLÈVES ?



Environ 30 % des professeurs de langue maternelle évaluent le travail des élèves par une note chiffrée synthétisant un certain nombre d'éléments. Ce mode d'évaluation est le plus souvent mentionné.

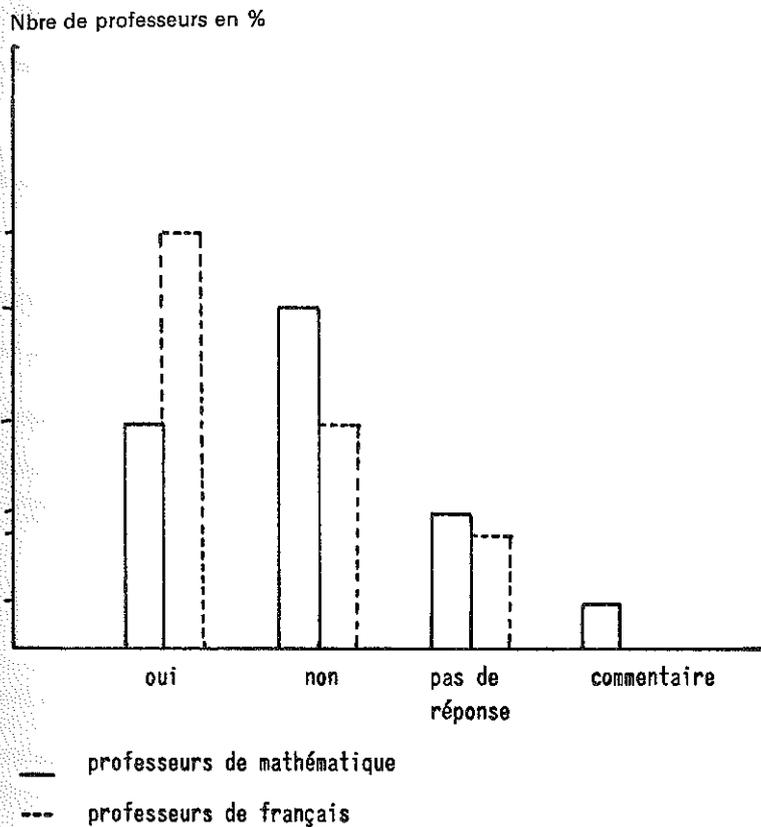


... et 63 % des professeurs de ...
 ... des élèves par une appréciation ...
 ... des pourcentages est inversée ...
 ... (1).

... professeurs utilisent simultanément ...
 ... globale. La comparaison des deux ...
 ... suppose que ces deux modes ...
 ... rit de certains enseignants.

- Une partie des professeurs de mathématique utilise exclusivement la note chiffrée.
- Une partie des professeurs de langue maternelle utilise exclusivement l'appréciation globale.

1.b.3. ÉVALUEZ-VOUS LE TRAVAIL D'UN ÉLÈVE PAR UNE NOTE SYNTHÉTISANT UN CERTAIN NOMBRE DE CRITÈRES CLAIREMENT DÉFINIS ?



Environ 30 % des professeurs de mathématique et 55 % des professeurs de langue maternelle évaluent le travail des élèves par une note synthétisant un certain nombre de critères bien définis. Les critères le plus souvent mentionnés ont été répartis en deux catégories.

ÉLÈVE PAR UNE APPRÉCIATION

pas de réponse commentaire

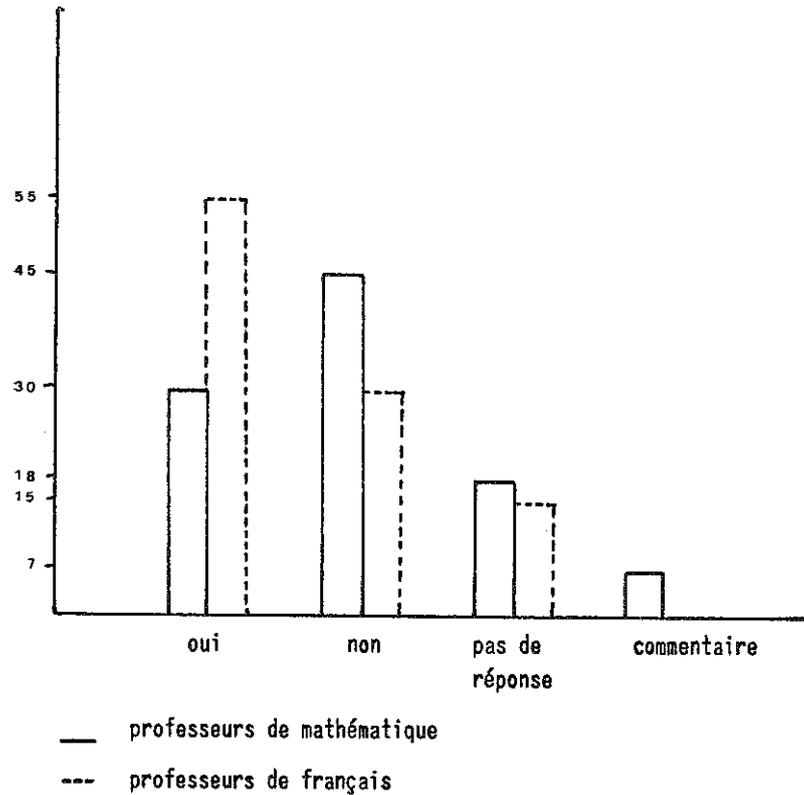
et 63 % des professeurs de élèves par une appréciation es pourcentages est inversée

es professeurs utilisent simultanément. La comparaison des deux modes apposer que ces deux modes de certains enseignants.

- . Une partie des professeurs de mathématique utilise exclusivement la note chiffrée.
- . Une partie des professeurs de langue maternelle utilise exclusivement l'appréciation globale.

1.b.3. ÉVALUEZ-VOUS LE TRAVAIL D'UN ÉLÈVE PAR UNE NOTE SYNTHÉTISANT UN CERTAIN NOMBRE DE CRITÈRES CLAIREMENT DÉFINIS ?

Nbre de professeurs en %



Environ 30 % des professeurs de mathématique et 55 % des professeurs de langue maternelle évaluent le travail des élèves par une note synthétisant un certain nombre de critères bien définis. Les critères le plus souvent mentionnés ont été répartis en deux catégories.

Langue maternelle.

1. Critères spécifiques à une tâche (surtout la rédaction).

Ex. : Fond, forme, orthographe, structure, richesse du vocabulaire, respect du sujet, enchaînement des idées, clarté de l'expression, etc.
52 professeurs.

2. Critères plus généraux, non spécifiques à une tâche.

Ex. : Application, esprit critique, effort, travail en classe, connaissances acquises, personnalité, progrès, initiative, « efficacité », netteté, soin, clarté, exactitude, rendement, etc.
57 professeurs.

Mathématique.

Les principaux critères proposés par les professeurs de mathématiques sont les suivants :

1. Méthode, exactitude, précision, forme de la présentation.

12 professeurs.

2. Capacité d'appliquer, aptitude au raisonnement, mémorisation, ordre, soin, activité, rapidité, etc.

60 professeurs.

De nombreux professeurs de mathématique insistent sur l'existence d'une seule mathématique, ce qui limiterait le nombre de critères spécifiques.

2. — FRÉQUENCE DES CONTRÔLES

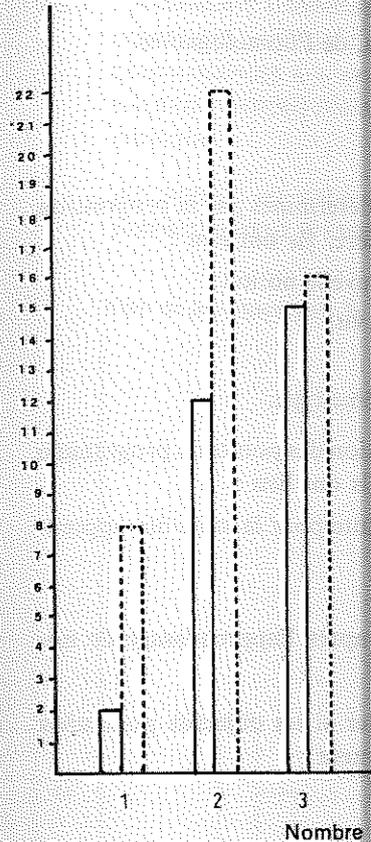
2.a. FRÉQUENCE DES CONTRÔLES ÉCRITS, PAR PÉRIODE ET PAR ÉLÈVE ?

Chez la majorité des professeurs, le nombre de contrôles écrits n'excède pas un par semaine.

— Mathématique : La fréquence des contrôles se situe principalement entre deux et six, par période et par élève.

— Langue maternelle : on observe un étalement plus grand (1 à 6). La fréquence la plus courante est de 2 à 3 contrôles par période (une période = six semaines).

Nbre de professeur en %



2.a.1. QUAND UN CONTRÔLE

Les commentaires faits par les professeurs sont répartis en trois catégories, la catégorie la plus fréquente est celle des contrôles écrits.

1. Raisons disciplinaires.

- Pour obliger les élèves à travailler.
- Vu la lenteur générale.
- Quand il y a un relâche.

out la rédaction).

re, richesse du vocabulaire, respect
clarté de l'expression, etc.

52 professeurs.

es à une tâche.

travail en classe, connaissances
initiative, « efficacité », netteté, soin,

57 professeurs.

s professeurs de mathématiques

e de la présentation.

12 professeurs.

nnement, mémorisation, ordre,

60 professeurs.

matique insistent sur l'existence
ait le nombre de critères spé-

CONTRÔLES

PAR PÉRIODE ET PAR ÉLÈVE ?

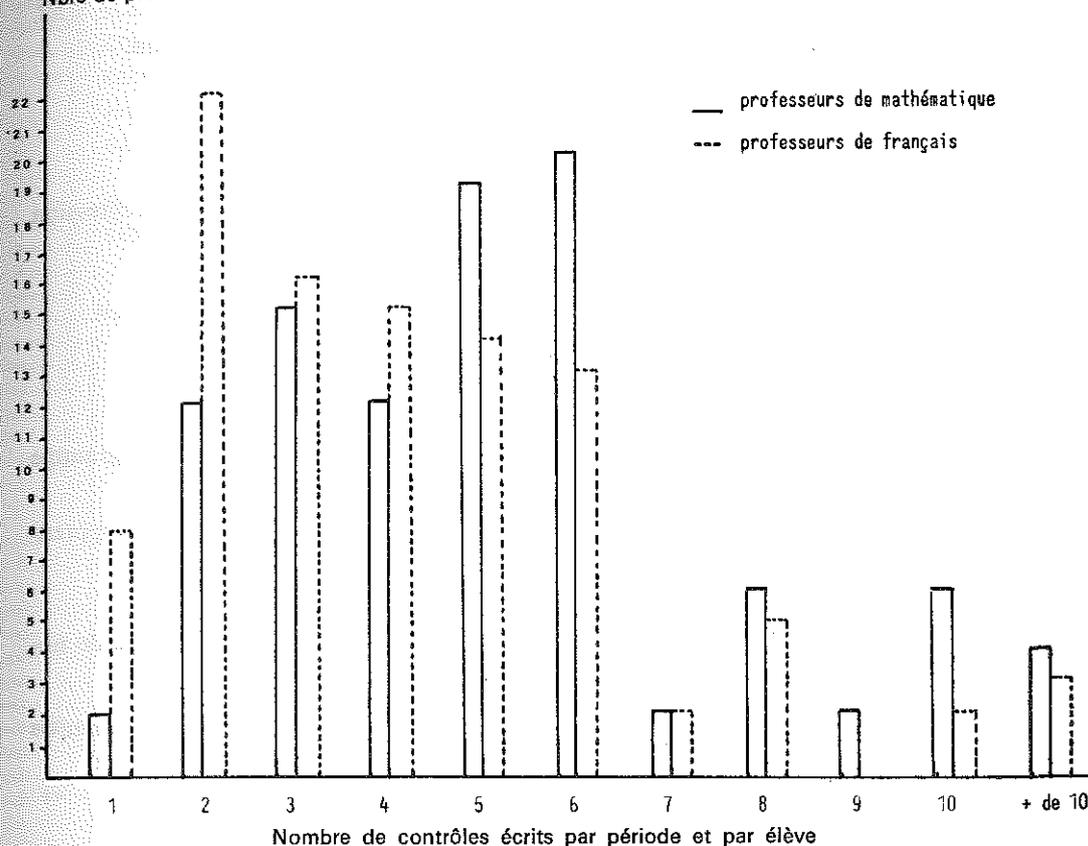
bre de contrôles écrits n'excède

trôles se situe principalement
ève.

talement plus grand (1 à 6).

2 à 3 contrôles par période

Nbre de professeur en %



2.a.1. QUAND UN CONTRÔLE ÉCRIT VOUS PARAÎT-IL NÉCESSAIRE ?

Les commentaires faits par les professeurs peuvent être répartis en trois catégories, la catégorie pédagogique comprenant cinq sous-catégories.

1. Raisons disciplinaires.

- . Pour obliger les élèves à travailler.
- . Vu la lenteur générale.
- . Quand il y a un relâchement.

2. Raisons administratives.

- . En raison de la nécessité d'un bulletin, de l'homologation.
- . En fin de période, de semaine.

3. Raisons pédagogiques.

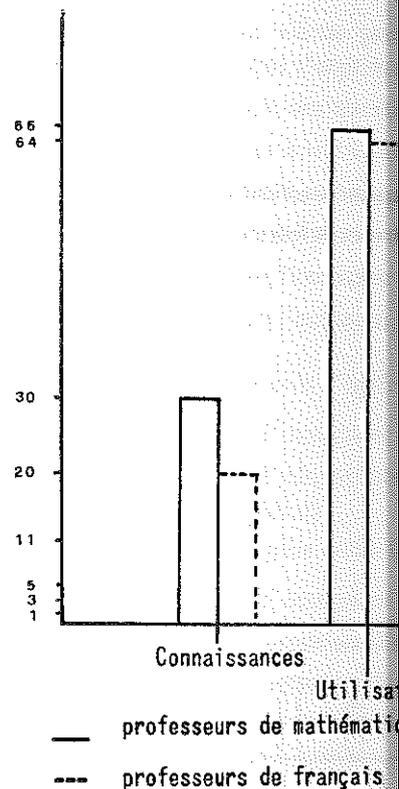
- a) Les contrôles sont *systématiquement* déterminés par la progression *logique* de l'enseignement de la matière. Ils ont lieu :
- . après l'enseignement de chaque nouvelle notion;
 - . quand un point du programme est vu;
 - . à la fin d'un chapitre;
 - . après une série d'exercices d'un même type.
- b) Les contrôles sont déterminés en fonction des points du programme enseignés et estimés fondamentaux. Ils ont lieu :
- . après l'étude d'une notion importante pour la suite des apprentissages;
 - . après une synthèse;
 - . après un ensemble d'exercices ou après l'apprentissage de notions formant un tout.
- c) Les contrôles ont pour but de vérifier si les objectifs poursuivis ont été atteints. On évalue principalement :
- . la connaissance théorique;
 - . la compréhension de la matière;
 - . la capacité d'appliquer.
- d) Souci d'observer certains comportements particuliers chez l'élève :
- . observer la fréquence d'apparition de certains types d'erreurs;
 - . observer l'acquisition d'une notion.
- e) Les contrôles s'inscrivent dans le cadre des exercices habituels :
- . orthographe, grammaire, rédaction (réponses propres aux professeurs de français);

80 % des professeurs de mathématique et 37 % des professeurs de langue maternelle justifient la nécessité d'un contrôle par des raisons pédagogiques (thème 3a).

2.a.2. IDÉALEMENT, L'ENSEIGNEMENT COMPAGNE, CONDUIT PROGRESSIVEMENT LES CONNAISSANCES À L'APPLICATION (INVENTION) POUR CULMINER EN UNE ÉVALUATION CRITIQUE.

AU NIVEAU OÙ VOUS ENSEIGNEZ, QUELLES SONT LES RAISONS PRINCIPALES QUI VOUS POUSSENT À UTILISER CE TYPE D'ÉVALUATION RESPECTIVEMENT À CES QUATRE NIVEAUX ?

Nbre de professeurs en %



Dans la hiérarchie idéale des objectifs, la majorité des professeurs (de mathématique) accordent la première place à l'évaluation.

tin, de l'homologation.

nt déterminés par la progression
atière. Ils ont lieu :

nouvelle notion;
est vu;

même type.

ction des points du programme
Ils ont lieu :

rtante pour la suite des appren-

ou après l'apprentissage de no-

rifier si les objectifs poursuivis
lement :

ments particuliers chez l'élève :
on de certains types d'erreurs;
on.

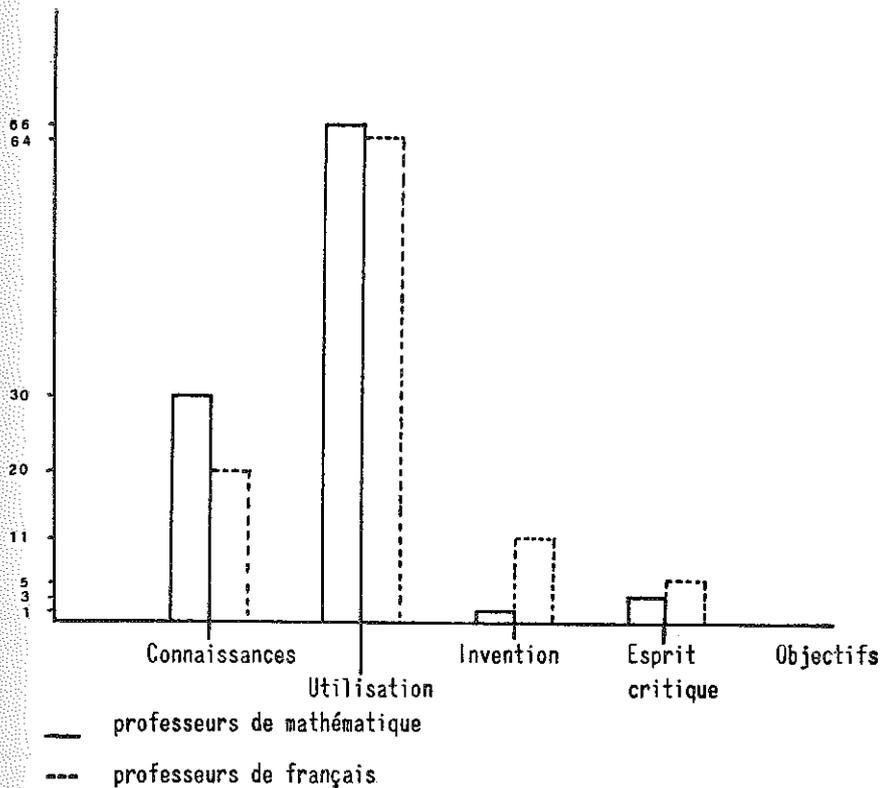
cadre des exercices habituels :
ction (réponses propres aux

e et 37 % des professeurs de
d'un contrôle par des raisons

2.a.2. IDÉALEMENT, L'ENSEIGNEMENT, ET DONC LE CONTRÔLE QUI L'ACCOMPAGNE, CONDUIT PROGRESSIVEMENT DE L'ACQUISITION DE CONNAISSANCES À L'APPLICATION, ENSUITE À LA RECHERCHE ORIGINALE (INVENTION) POUR CULMINER ENFIN AVEC L'USAGE DE L'ESPRIT CRITIQUE.

AU NIVEAU OÙ VOUS ENSEIGNEZ QUELLE PLACE ACCORDEZ-VOUS RESPECTIVEMENT À CES QUATRE OBJECTIFS ?

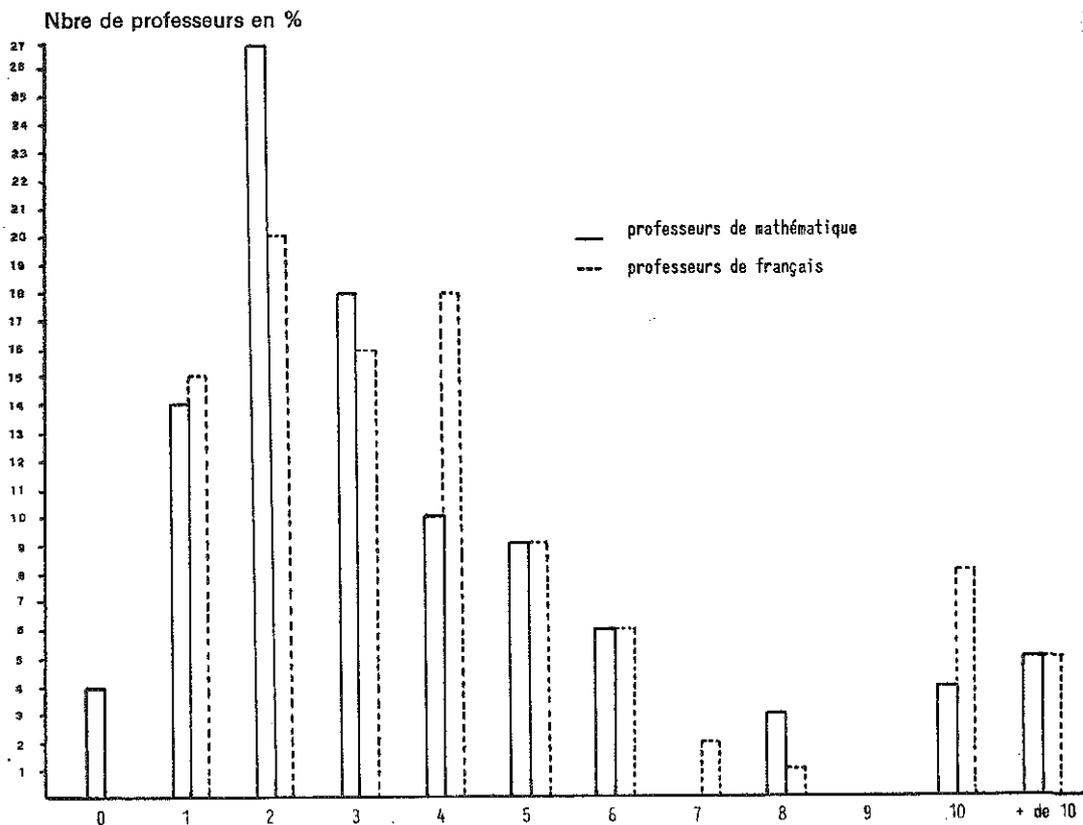
Nbre de professeurs en %



Dans la hiérarchie idéale des objectifs sur lesquels porte l'évaluation, la majorité des professeurs (de mathématique et de langue maternelle) accordent la première place à l'utilisation des connaissances (environ

65 % chez les deux). Ensuite vient la capacité de restituer la matière (mathématique : 30 %; langue maternelle : 20 %). Pour la mathématique, la place réservée à l'invention est négligeable (1 %). Elle n'est d'ailleurs pas très élevée en langue maternelle non plus (11 %), alors qu'ici les occasions restent cependant beaucoup plus nombreuses.

2.b. FRÉQUENCE DES CONTRÔLES ORAUX, PAR PÉRIODE ET PAR ÉLÈVE ?



2.b.1. QUE VOUS APPORTE LE CONTRÔLE ÉCRIT ?

Trois grandes catégories perçues par les professeurs :

1. Réponses envisagées sous l'angle de l'élève

- 1.1. Evaluation immédiate.
- 1.2. Aide pour éviter de nou...
- 1.3. Encouragement pour les...
- 1.4. Favorable aux élèves p...
ou aux élèves qui s'expr...
- 1.5. Individualisation des qu...
- 1.6. Oblige les élèves à trava...
- 1.7. Donne à l'élève le sentin...
- 1.8. Valorise l'élève devant l...
- 1.9. Donne à l'élève l'ocasi...
- 1.10. Donne à l'élève l'ocasi...

2. Réponses envisagées sous l'angle de l'enseignant

2.a. Dans son rôle d'enseignant

- 2.a.1. Permet d'adapter l'enseignant (est immédiat).
- 2.a.2. Contact plus facile.

2.b. Dans son rôle d'enseignant

- 2.b.1. Est une occasion de...
- 2.b.2. Est une occasion de...
- 2.b.3. Est une occasion de...
(rappel ou synthèse)

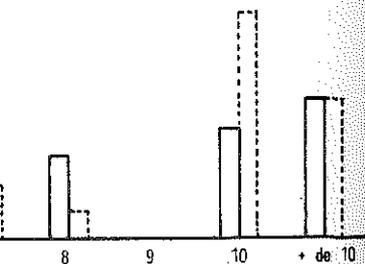
2.c. Dans son rôle de contrôleur

- 2.c.1. Contrôle direct
- 2.c.2. Rapidité (plus r...
- 2.c.3. Permet une plus...

capacité de restituer la matière (20 %). Pour la mathématique, elle est de 11 %, et pour le français de 10 % (11 %). Elle n'est d'ailleurs pas la même pour tous les élèves (11 %), alors qu'ici les résultats sont nombreux.

PAR PÉRIODE ET PAR ÉLÈVE ?

de mathématique
de français



2.b.1. QUE VOUS APPORTE LE CONTRÔLE ORAL PAR RAPPORT AU CONTRÔLE ÉCRIT ?

Trois grandes catégories permettent de regrouper les commentaires des professeurs :

1. Réponses envisagées sous l'angle de l'avantage pour l'élève.

- 1.1. Evaluation immédiate.
- 1.2. Aide pour éviter de nouvelles erreurs, pour remettre en piste.
- 1.3. Encouragement pour les élèves déficients.
- 1.4. Favorable aux élèves pris de panique aux épreuves écrites, ou aux élèves qui s'expriment difficilement par écrit.
- 1.5. Individualisation des questions selon les capacités.
- 1.6. Oblige les élèves à travailler régulièrement.
- 1.7. Donne à l'élève le sentiment d'être suivi régulièrement.
- 1.8. Valorise l'élève devant la classe.
- 1.9. Donne à l'élève l'occasion d'exercer son esprit critique
- 1.10. Donne à l'élève l'occasion de s'exprimer.

2. Réponses envisagées sous l'angle de l'avantage pour le maître.

2.a. Dans son rôle d'enseignant en général.

- 2.a.1. Permet d'adapter son enseignement (parce que le contrôle est immédiat).
- 2.a.2. Contact plus facile (dialogue).

2.b. Dans son rôle d'enseignant de manière spécifique.

- 2.b.1. Est une occasion d'exercice d'élocution.
- 2.b.2. Est une occasion de corriger les erreurs de langage.
- 2.b.3. Est une occasion de rafraîchir la mémoire de la classe (rappel ou synthèse de leçons antérieures).

2.c. Dans son rôle de contrôleur.

- 2.c.1. Contrôle direct de la matière en cours.
- 2.c.2. Rapidité (plus rapide que l'écrit).
- 2.c.3. Permet une plus grande diversité des types d'évaluation.

2.c.4. Plus profond, plus nuancé, plus objectif que l'écrit.
Confirme l'écrit.

3. *L'oral est plus efficace que l'écrit pour évaluer certains aspects.*

- 3.1. Esprit de synthèse, compréhension et utilisation d'une notion.
- 3.2. Rapidité de résolution d'un problème.
- 3.3. Spontanéité, esprit de répartie, créativité.
- 3.4. Autres aspects :
 - 3.4.a. Permet d'observer les réactions de l'élève devant une difficulté.
 - 3.4.b. Fournit une meilleure connaissance de l'élève.
 - 3.4.c. Permet d'observer la personnalité de l'élève.

3. — SYNTHÈSE DES ÉVALUATIONS EN FIN DE PÉRIODE

3.a. COMMENT SYNTHÉTISEZ-VOUS VOS ÉVALUATIONS EN FIN DE PÉRIODE ?

Les réponses sont classées d'après les commentaires faits par les professeurs. On distingue d'abord deux grandes catégories :

- A. Une catégorie générale regroupant les réponses des professeurs qui restent à un niveau qualitatif :
 - . Utilisation des graphiques;
 - . Observation de la tendance dominante (en tenant compte ou non des notes extrêmes).
- B. Une catégorie générale regroupant les réponses des professeurs qui passent par une traduction en chiffres : calcul de la moyenne des notes chiffrées et transposition de cette moyenne dans la nouvelle échelle.

Trois sous-catégories ont, en outre, été introduites :

- a) Référence au groupe classe (calcul de la moyenne générale de la classe et ajustement des moyennes individuelles).
- b) Prise en considération des efforts, des progrès, des capacités de l'élève.
- c) Mention d'évaluations partielles portant sur différentes composantes

d'une discipline (résultats de l'écrit, etc.) qui précèdent l'établissement de la période).

La réponse d'un professeur porte sur les éléments qu'elle concerne.

Enfin, on a distingué une troisième catégorie de réponses mentionnant explicitement le calcul de l'écart-type.

Tendance générale.

De nombreux professeurs ont mentionné les efforts, les progrès, les capacités de l'élève. Le degré de l'échelle à cinq niveaux sera confirmée lors de l'analyse des données de l'élève intervenant dans la détermination du degré atteint par l'élève.

Recours à la statistique (calcul de la moyenne arithmétique). Il est plus fréquent dans les réponses de mathématiques que par ailleurs.

La question 1.b.1. portait sur le fait que 62 % des professeurs de mathématiques ont déclaré y recourir.

Pour la synthèse des résultats de mathématiques et de physique, on a obtenu : 91 professeurs de mathématiques.

Il semble que des professeurs ont mentionné des chiffres lors de l'appréciation de l'élève au moment de la synthèse.

3.b. QUELLES DIFFICULTÉS RENCONTREZ-VOUS ?

Les deux principales sources de difficultés sont :

é, plus objectif que l'écrit.
ur évaluer certains aspects.
on et utilisation d'une notion.
blème.
éativité.

actions de l'élève devant une
naissance de l'élève.
nnalité de l'élève.

NS EN FIN DE PÉRIODE

LUATIONS EN FIN DE PÉRIODE ?

s commentaires faits par les
grandes catégories :

les réponses des professeurs

nte (en tenant compte ou non

les réponses des professeurs
iffres : calcul de la moyenne
ette moyenne dans la nouvelle

introduites :

e la moyenne générale de la
ividuelles).

ogrès, des capacités de l'élève.

t sur différentes composantes

d'une discipline (résultats de la période en orthographe, en rédaction, etc.) qui précèdent l'établissement de la moyenne des résultats de la période).

La réponse d'un professeur peut être classée dans plusieurs catégories, selon les éléments qu'elle contient.

Enfin, on a distingué une troisième grande catégorie (C) qui regroupe les réponses mentionnant explicitement le recours à la statistique (principalement calcul de l'écart-type et courbe de Gauss).

Tendance générale.

De nombreux professeurs (80 professeurs de mathématique et 83 professeurs de langue maternelle) déclarent prendre en considération les efforts, les progrès, les capacités des élèves pour les situer sur un degré de l'échelle à cinq niveaux (catégories Ab et Bb). Cette tendance sera confirmée lors de l'analyse de la question 3.c.1. (Les qualités humaines de l'élève interviennent dans la note en entraînant une modification du degré atteint par l'élève dans l'échelle).

Recours à la statistique (calcul du sigma, courbe de Gauss, moyenne arithmétique). Il est plus fréquemment signalé par les professeurs de mathématique que par leurs collègues enseignant le français.

La question 1.b.1. portait déjà sur l'utilisation de notes chiffrées : 62 % des professeurs de mathématique et 50 % des professeurs de français ont déclaré y recourir.

Pour la synthèse des résultats en fin de période, l'écart entre les professeurs de mathématique et les professeurs de français s'accroît nettement : 91 professeurs de mathématique et 44 professeurs de français.

Il semble que des professeurs de français utilisant des notes numériques lors de l'appréciation d'une performance isolée n'ont plus recours à des chiffres au moment de la synthèse générale.

3.b. QUELLES DIFFICULTÉS RENCONTREZ-VOUS ?

Les deux principales sources de difficultés résident en deux points :

1. La synthèse des résultats et la traduction par un niveau de l'échelle : 32 professeurs de mathématique et 40 professeurs de français.
2. Le trop petit nombre de degrés de l'échelle à cinq degrés, son manque de nuances : 33 professeurs de mathématique et 30 professeurs de français.

De nombreux professeurs déclarent n'éprouver aucune difficulté. D'autres formulent un certain nombre de griefs; ils ont été répartis en cinq catégories :

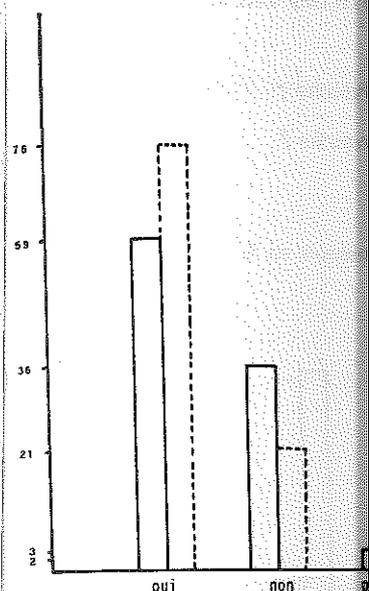
- a) *Concernant surtout l'utilisation de chiffres.*
 - . Manque de connaissance de la statistique descriptive.
 - . Difficultés de calcul.
- b) *En relation avec l'évaluation qualitative.*
 - . Difficulté de la synthèse.
Exemple : TB pour le raisonnement + F pour la mémorisation = F, S ou ?
 - . Difficulté de traduire la tendance générale d'un élève.
 - . Difficulté d'être objectif.
- c) *Manque de temps.*
Fluctuation du comportement des élèves.
- d) *Difficulté d'être objectif en ne décourageant pas l'élève.*
Trop peu de degrés dans l'échelle.
- e) *Par rapport à la classe.*
 - . Difficulté de cerner la valeur de l'élève en soi (l'effort de l'élève peut être perdu dans l'effort de l'ensemble de la classe).
 - . Difficulté d'évaluer l'élève par rapport à la moyenne de la classe.

3.c. LES QUALITÉS HUMAINES DE L'ÉLÈVE INTERVIENNENT DANS LA NOTE :

3.c.1. *en entraînant parfois une modification du degré atteint par l'élève dans l'échelle.*

3.c.2. *sous la forme d'un commentaire joint à l'évaluation.*

Nbre de professeurs en %



59 % des professeurs de langue maternelle disent m humaines de l'élève.

94 % des professeurs de intervenir celles-ci sous la fo

4. — Co

4.a.1. VOS COLLÈGUES EN UTILISENT-ILS LES MÊM

70 % des professeurs de utiliser les mêmes types de

on par un niveau de l'échelle :
0 professeurs de français.

nelle à cinq degrés, son manque
ématique et 30 professeurs de

n'éprouver aucune difficulté.
e griefs; ils ont été répartis en

iffres.

istique descriptive.

+ F pour la mémorisation =

générale d'un élève.

es.

geant pas l'élève.

élève en soi (l'effort de l'élève
semble de la classe).

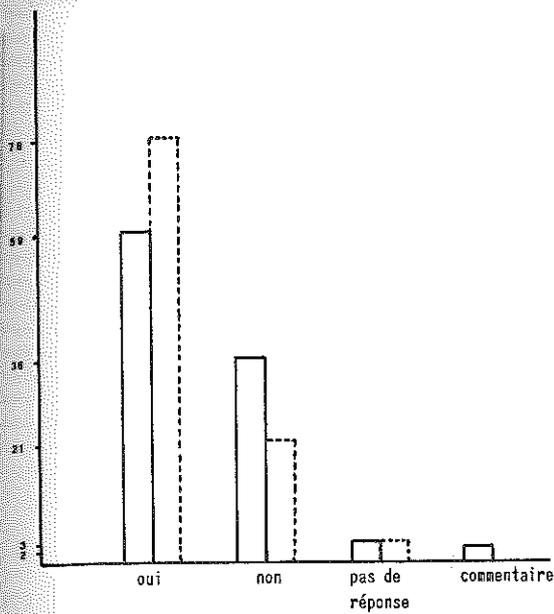
ort à la moyenne de la classe.

INTERVIENNENT DANS LA NOTE :

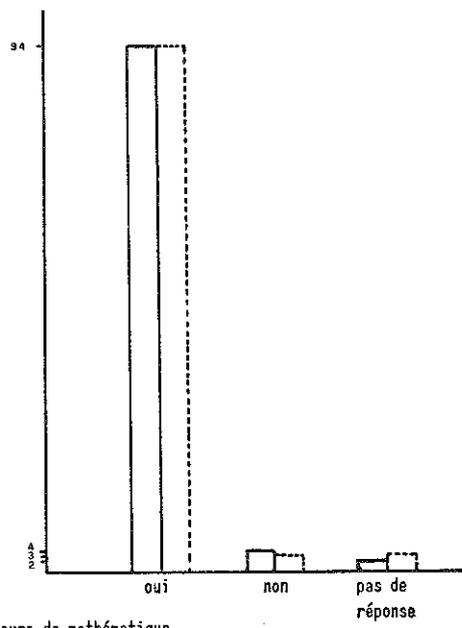
ification du degré atteint par

joint à l'évaluation.

Nbre de professeurs en %



Nbre de professeurs en %



— professeurs de mathématique

- - - professeurs de français

59 % des professeurs de mathématique et 76 % des professeurs de langue maternelle disent modifier leur note en fonction des qualités humaines de l'élève.

94 % des professeurs de mathématique et de langue maternelle font intervenir celles-ci sous la forme d'un commentaire joint à l'évaluation.

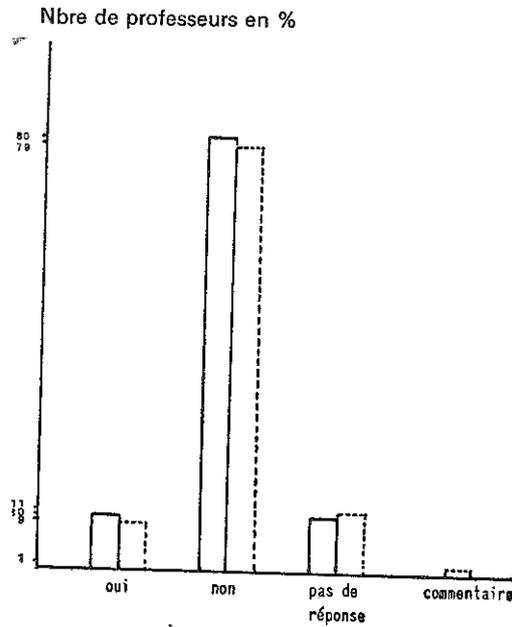
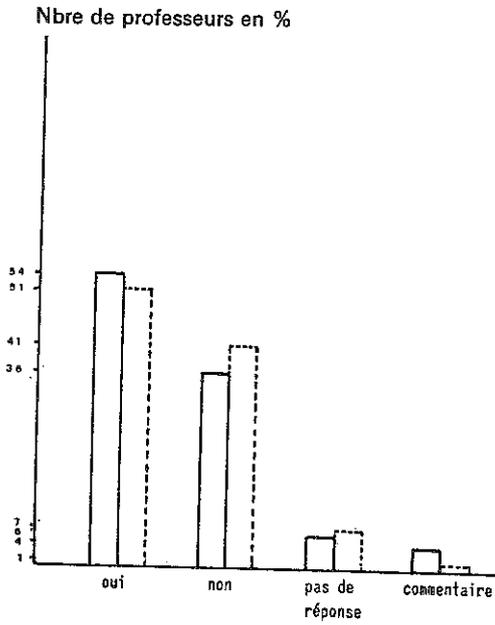
4. — COORDINATION DES CONTRÔLES

4.a.1. VOS COLLÈGUES ENSEIGNANT LA MÊME DISCIPLINE QUE VOUS UTILISENT-ILS LES MÊMES TYPES DE CONTRÔLE QUE VOUS ?

70 % des professeurs de mathématique *et* de langue maternelle disent utiliser les mêmes types de contrôle que leurs collègues.

4.a.2. LES TYPES DE CONTRÔLE QUE VOUS UTILISEZ ONT-ILS ÉTÉ ÉLABORÉS EN COLLABORATION:

- a) avec des professeurs de la même branche ;
- b) avec des professeurs d'une autre branche ?

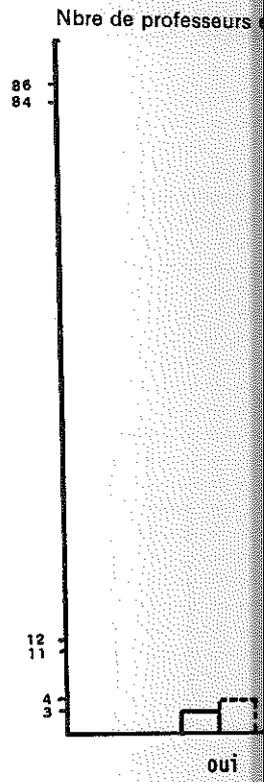


Collaboration avec des collègues enseignant la même branche

Collaboration avec des collègues enseignant une autre branche

— professeurs de mathématique
 --- professeurs de français

c) avec des professeurs d'



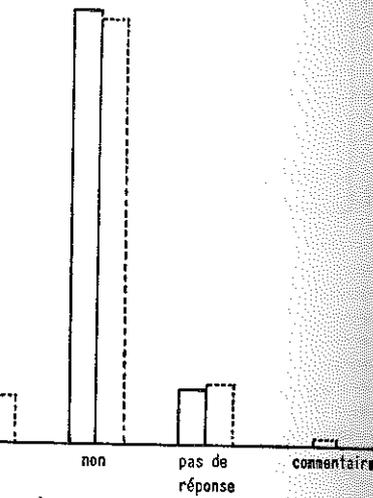
— professeurs
 --- professeurs

S UTILISEZ ONT-ILS ÉTÉ ÉLABORÉS

e branche ;

e branche ?

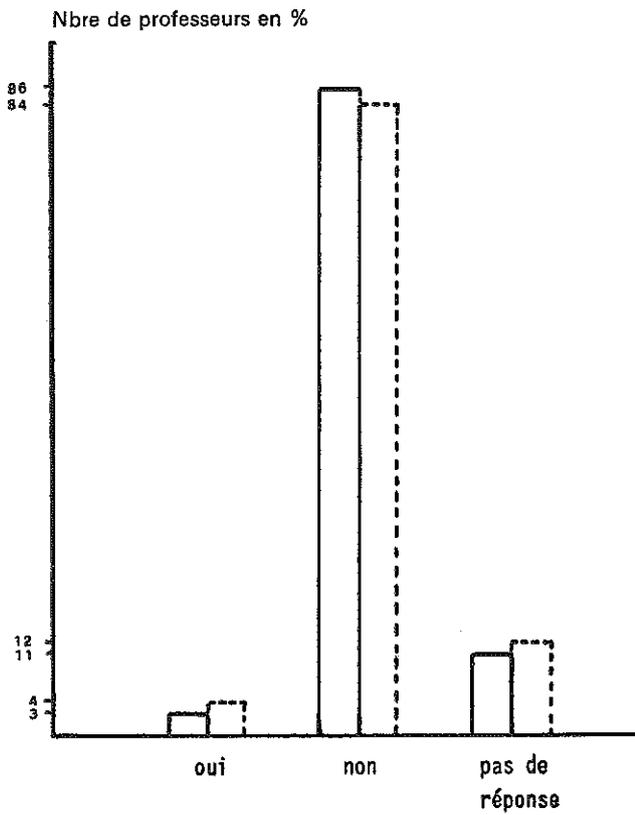
professeurs en %



tion avec des collègues enseignant
une autre branche

ue

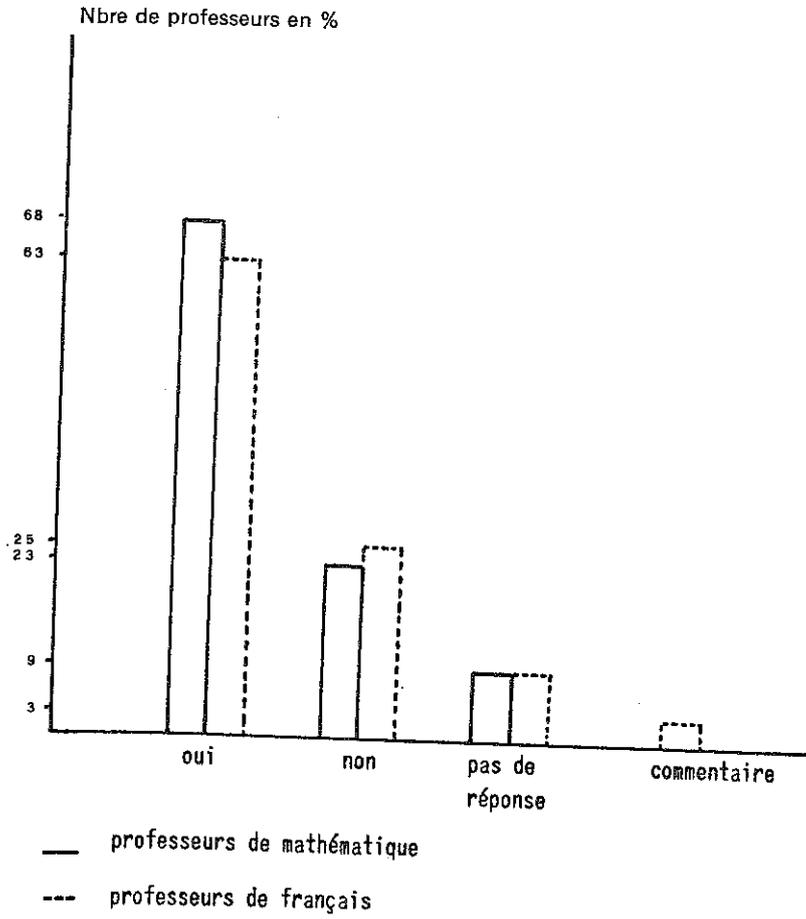
c) avec des professeurs d'un autre établissement.



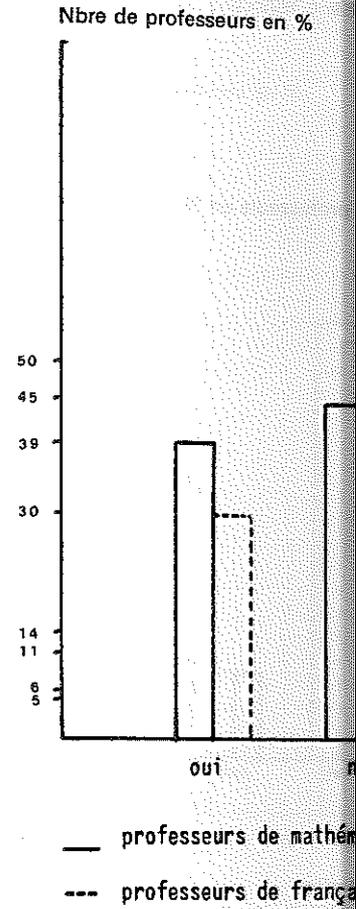
— professeurs de mathématique
- - - professeurs de français

4.b. VOYEZ-VOUS LA POSSIBILITÉ D'ADOPTER DES CRITÈRES D'ÉVALUATION COMMUNS :

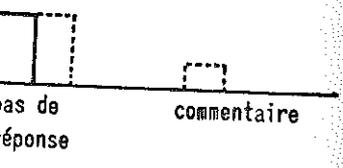
a) à l'ensemble des professeurs de langue maternelle/de mathématique d'un même établissement ?



b) à plusieurs établissements

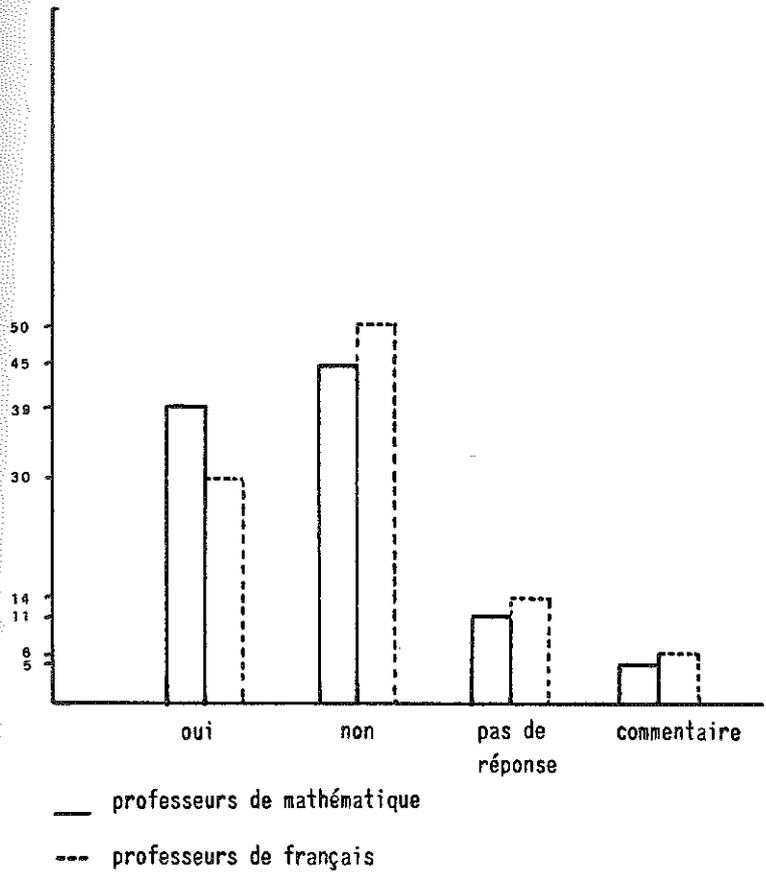


langue maternelle/de mathéma-



b) à plusieurs établissements ?

Nbre de professeurs en %

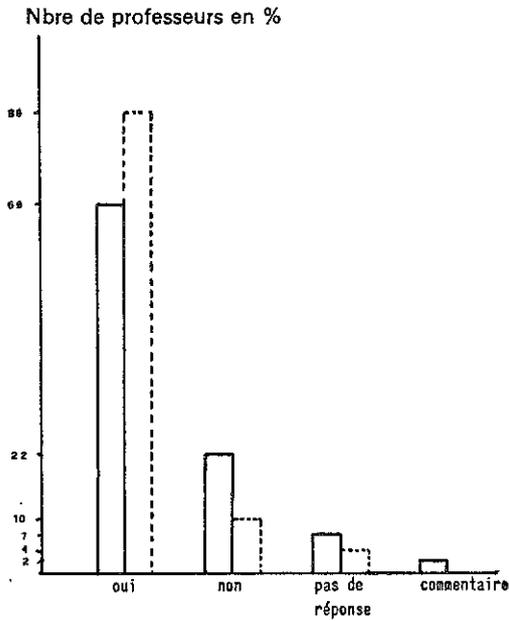


5. — NORMES

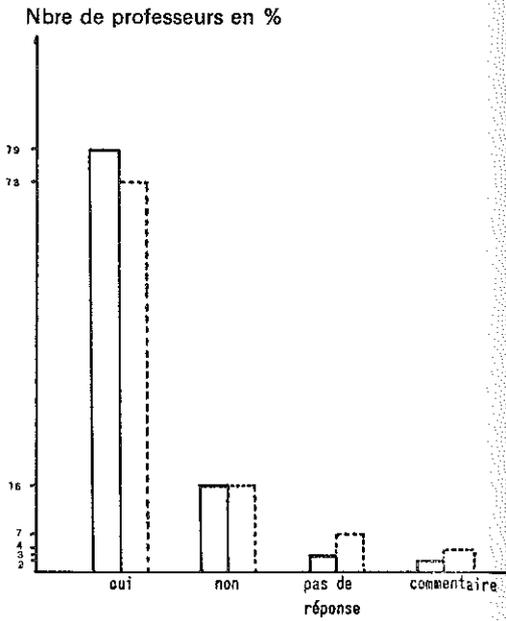
5.a. Vos APPRÉCIATIONS VISENT À SITUER L'ÉLÈVE :

5.a.1. *par rapport à lui-même.*

5.a.2. *par rapport à sa classe.*



Situer l'élève par rapport à lui-même

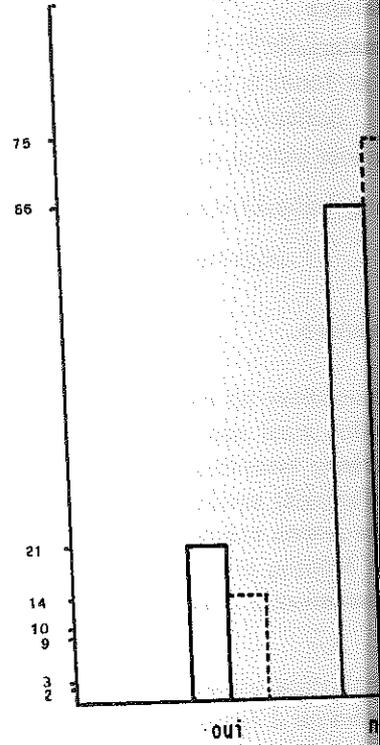


Situer l'élève par rapport à sa classe

— professeurs de mathématiques
 --- professeurs de français

5.a.3. *par rapport à une population des classes de premier degré régional, provinciale...*

Nbre de professeurs en %



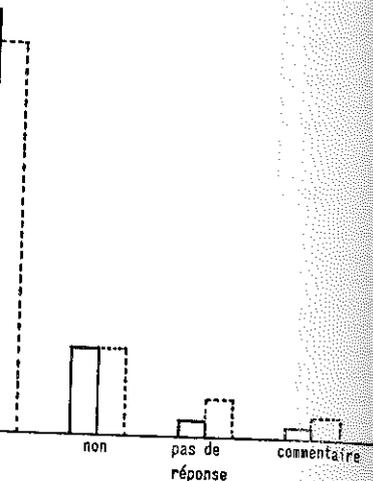
— professeurs de mathématiques
 --- professeurs de français

Situation de l'élève par rapport à...

69 % des professeurs de mathématiques
 86 % des professeurs de français

ER L'ÉLÈVE :

professeurs en %

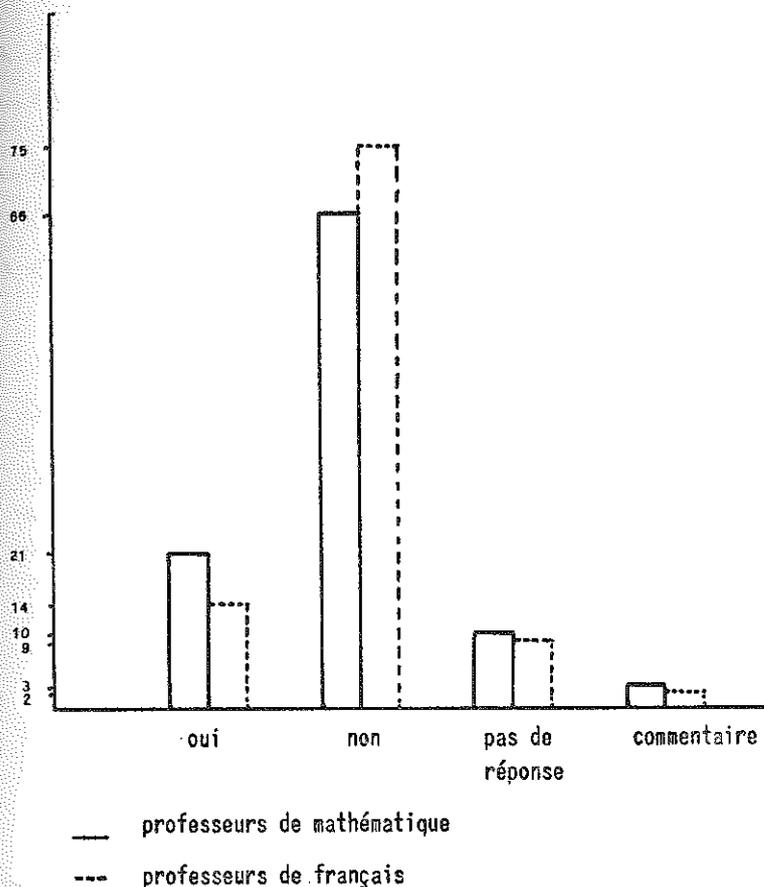


l'élève par rapport à sa classe

ique

5.a.3. par rapport à une population plus grande (exemples : ensemble des classes de première de l'enseignement rénové; éta-lonnage régional, provincial,...).

Nbre de professeurs en %



Situation de l'élève par rapport à lui-même :

69 % des professeurs de mathématique

86 % des professeurs de langue maternelle.

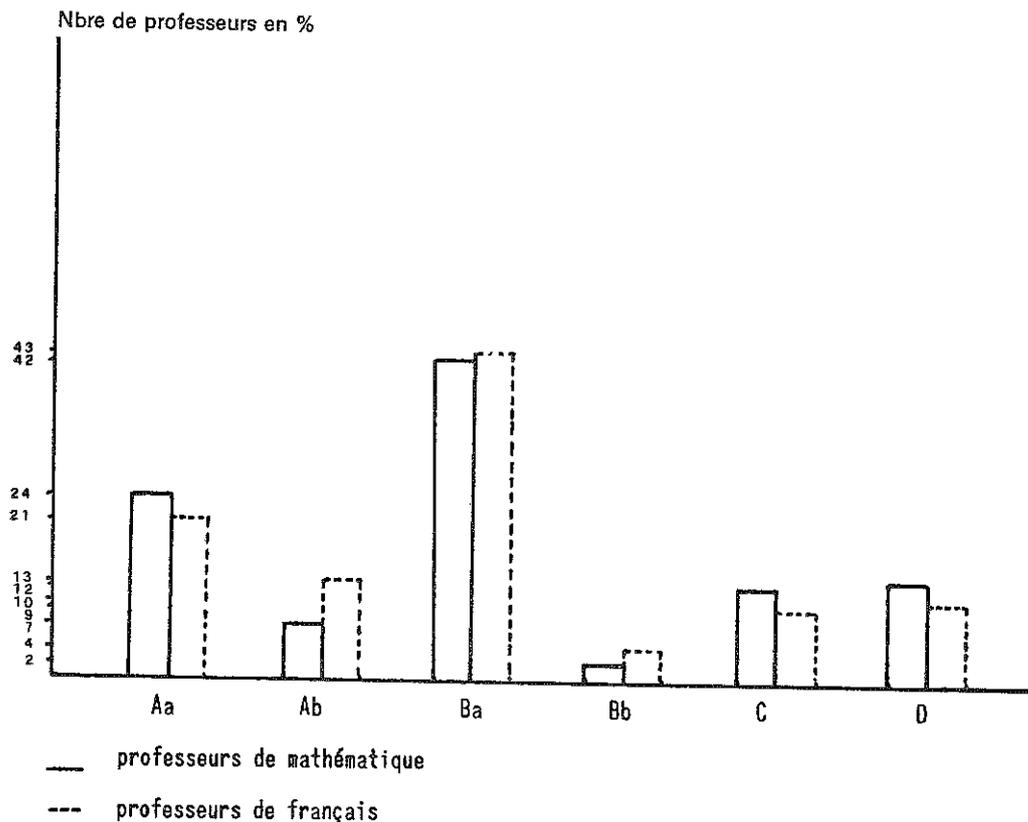
Situation de l'élève par rapport à la classe :

- 79 % des professeurs de mathématique
- 73 % des professeurs de langue maternelle.

Situation de l'élève par rapport à une population plus grande :

- 21 % des professeurs de mathématique
- 14 % des professeurs de langue maternelle.

5.b. COMMENT PROCÉDEZ-VOUS ?



On distingue six types de réponses :

- A. Par rapport à lui-même :
 - a) en comparant le résultat à la moyenne individuelle;
 - b) en lisant le graphique de la courbe de Gauss.
- B. Par rapport à la classe :
 - a) en établissant d'abord la moyenne individuelle;
 - b) en cherchant la note moyenne de la classe.
- C. Courbe de Gauss (dessin à l'échelle).
- D. Calcul de l'écart-type.

Analyses

1. Les professeurs de mathématique ont répondu à la question que leurs collègues de français ont posée. Ces derniers indiquent généralement un groupe d'élèves (Exemple : les classes de première année; etc.). Toutefois, le procédé de comparaison de l'élève par rapport à lui-même n'est pas indiqué.
 2. Les deux procédés généraux de comparaison de l'élève par rapport à lui-même sont mentionnés dans l'ordre :
 - . La comparaison du résultat à la moyenne individuelle.
 - . L'ajustement des moyennes individuelles à la moyenne de la classe.
 3. Le recours à la statistique (calcul de l'écart-type) est rarement mentionné.
- Il ressort donc de cette analyse que le recours à l'élève, tant par rapport à lui-même qu'à la moyenne de la classe, est encore peu répandu.

: :
ne
nelle.
ulation plus grande :
e
nelle.

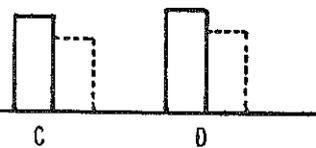
On distingue six types de réponses :

- A. *Par rapport à lui-même* :
- a) en comparant le résultat actuel aux résultats précédents;
 - b) en lisant le graphique des résultats enregistrés au cours de la période.
- B. *Par rapport à la classe* :
- a) en établissant d'abord la moyenne de la classe et ensuite les moyennes individuelles;
 - b) en cherchant la note modale et la position dans le groupe.
- C. Courbe de Gauss (dessin à partir de l'histogramme).
- D. Calcul de l'écart-type.

Analyse des résultats

1. Les professeurs de mathématique répondent de façon plus précise à la question que leurs collègues enseignant la langue maternelle. Ces derniers indiquent généralement le moyen utilisé pour évaluer un groupe d'élèves (Exemples : exercice proposé en commun aux classes de première année; test collectif élaboré par les professeurs, etc.). Toutefois, le procédé mathématique qui permet de situer un élève par rapport à lui-même et par rapport à la classe n'est jamais indiqué.
2. Les deux procédés généralement utilisés pour situer respectivement l'élève par rapport à lui-même et par rapport à la classe sont, dans l'ordre :
 - . La comparaison du résultat actuel aux résultats précédents;
 - . L'ajustement des moyennes individuelles en fonction de la moyenne de la classe.
3. Le recours à la statistique (courbe de Gauss et calcul de l'écart-type) est rarement mentionné.

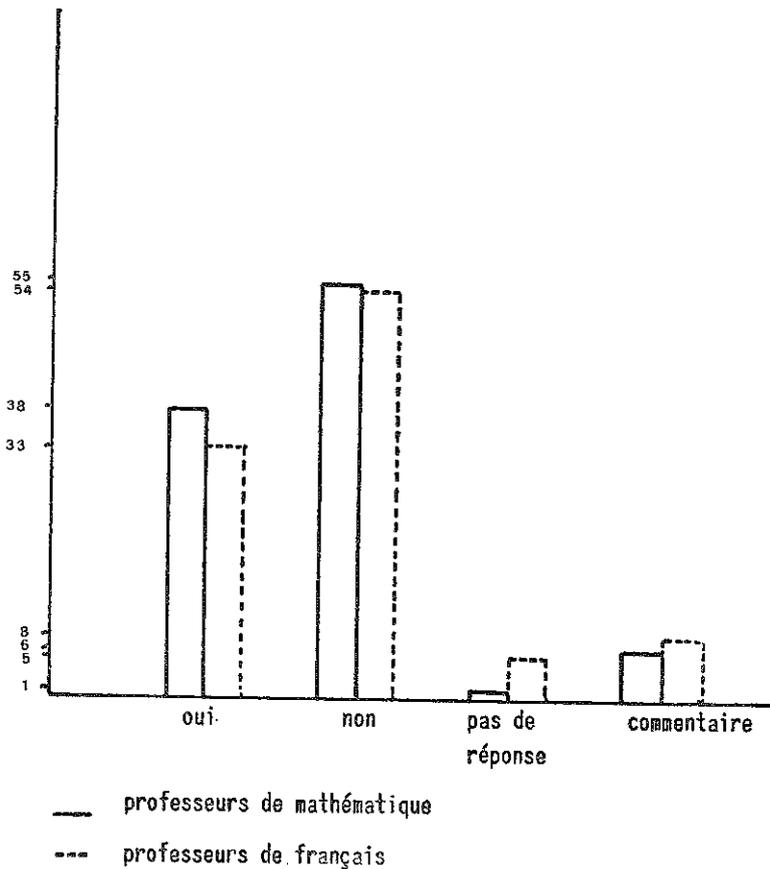
Il ressort donc de cette analyse des procédés utilisés pour situer un élève, tant par rapport à lui-même que par rapport à la classe, que l'hétérogénéité des méthodes est encore grande.



6. — DIFFICULTÉS RENCONTRÉES

6.a.1. LE NOUVEAU SYSTÈME DE CONTRÔLE VOUS PARAÎT-IL MOINS PRÉCIS QUE L'ANCIEN ? EN QUOI ?

Nbre de professeurs en %



Le nouveau système est moins précis que l'ancien :

38 % des professeurs de mathématique

33 % des professeurs de français.

Le nouveau système n'est pas mo

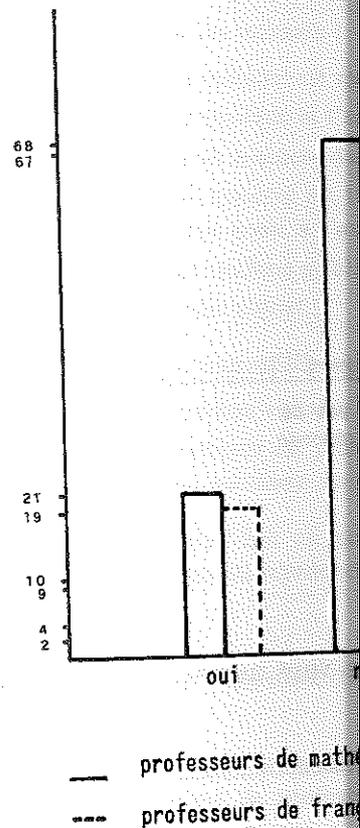
55 % des professeurs de math

54 % des professeurs de fran

Les raisons données par les pr
veau système est moins précis q
à la question 1.a.2. Nombre de
nuances de l'échelle à cinq degré
52 professeurs de langue matern

6.a.2. LE NOUVEAU SYSTÈME DE C
QUE L'ANCIEN ?

Nbre de professeurs en %



— professeurs de math

--- professeurs de fran

ENCONTRÉES

LE VOUS PARAÎT-IL MOINS PRÉCIS

Le nouveau système n'est pas moins précis que l'ancien :

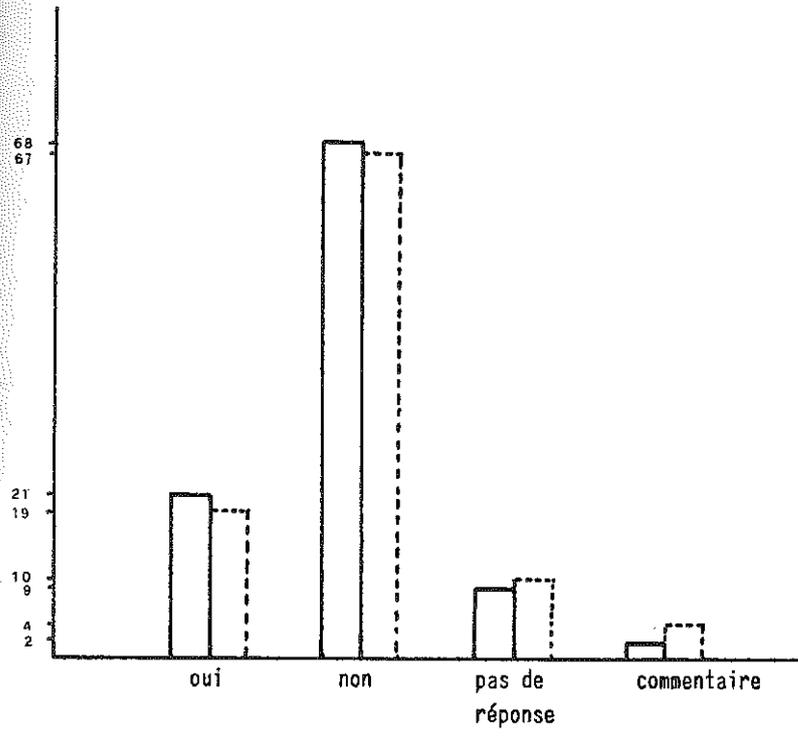
55 % des professeurs de mathématique

54 % des professeurs de français.

Les raisons données par les professeurs qui considèrent que le nouveau système est moins précis que l'ancien rejoignent notre remarque à la question 1.a.2. Nombre de maîtres se plaignant du manque de nuances de l'échelle à cinq degrés : 58 professeurs de mathématique et 52 professeurs de langue maternelle.

6.a.2. LE NOUVEAU SYSTÈME DE CONTRÔLE VOUS PARAÎT-IL MOINS OBJECTIF QUE L'ANCIEN ?

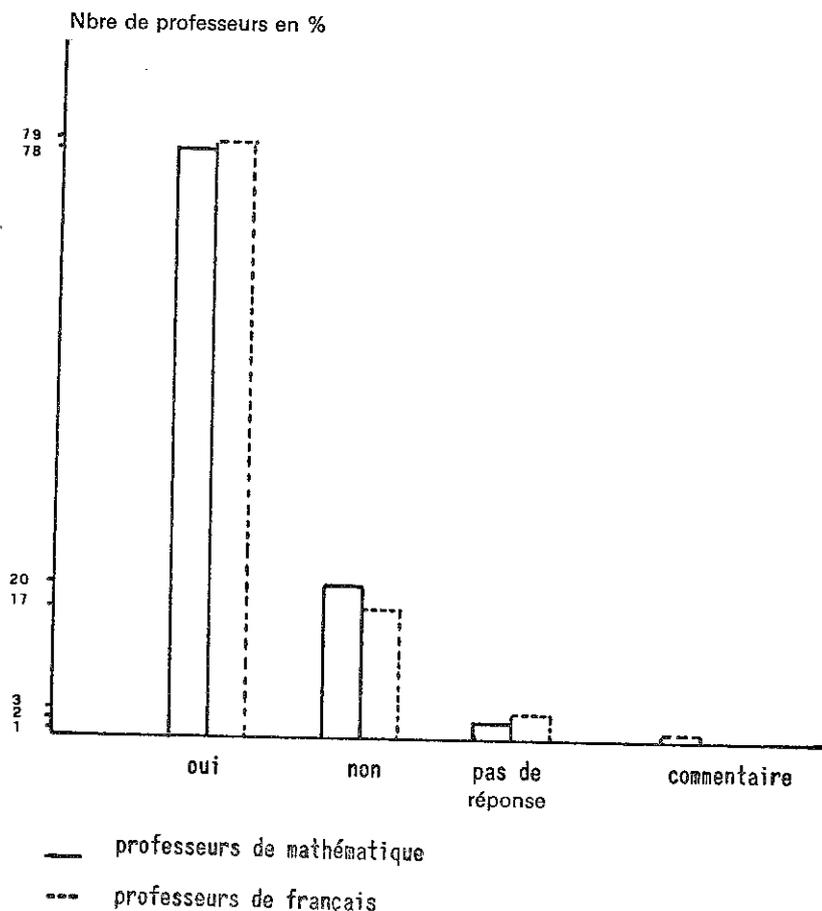
Nbre de professeurs en %



— professeurs de mathématique

- - - professeurs de français

6.b. LE NOUVEAU SYSTÈME DE CONTRÔLE VOUS PREND-IL PLUS DE TEMPS QUE L'ANCIEN ?



6.c. LE NOUVEAU SYSTÈME DE CONTRÔLE VOUS SATISFAIT-IL PLUS QUE L'ANCIEN ? EN QUOI ?

Distinguons d'abord le groupe des professeurs satisfaits du groupe des professeurs non satisfaits :

- I. Les points sur lesquels porte
 - a) Le nouveau système est p
 - b) Le nouveau système est p
 - c) Le nouveau système perm
 - d) Il permet de mieux situer
 - e) Il permet la comparaiso
 - d'une autre classe.
 - f) Il permet une appréciatio
 - g) Le nouveau système est p
 - h) Il est moins contraignant
 - i) Il est plus précis, plus ob
 - sance de l'élève.
- II. Les griefs sont les suivants :
 - a) Le nouveau système est r
 - b) Le nouveau système et l'
 - c) Le manque de nuances d
 - d) Le nouveau système est
 - l'autre, d'une école à l'
 - e) Le nouveau système oub
 - f) L'élève étudie moins.

Les principales sources de sa

1. Le nouveau système est *plus*
2. Le nouveau système est plu
- nuancée et une *meilleure co*

Le principal grief fait au n

mathématiques qui lui reproch

7. — REMARQUES COMPLÉMENT

Voici d'abord, par ordre de

nombre assez élevé de profes

VOUS PREND-IL PLUS DE TEMPS

I. Les points sur lesquels porte la satisfaction sont les suivants :

- a) Le nouveau système est plus humain.
- b) Le nouveau système est plus sécurisant pour l'élève.
- c) Le nouveau système permet de mieux suivre l'élève.
- d) Il permet de mieux situer l'élève par rapport à lui-même.
- e) Il permet la comparaison des résultats d'une classe avec ceux d'une autre classe.
- f) Il permet une appréciation plus nuancée.
- g) Le nouveau système est plus logique.
- h) Il est moins contraignant, moins aride.
- i) Il est plus précis, plus objectif, et apporte une meilleure connaissance de l'élève.

II. Les griefs sont les suivants :

- a) Le nouveau système est moins précis.
- b) Le nouveau système et l'ancien sont de valeur égale.
- c) Le manque de nuances dans la notation décourage l'élève.
- d) Le nouveau système est trop subjectif (il varie d'un professeur à l'autre, d'une école à l'autre).
- e) Le nouveau système oublie l'émulation.
- f) L'élève étudie moins.

Les principales sources de satisfaction sont dans l'ordre :

1. Le nouveau système est *plus humain*.
2. Le nouveau système est plus objectif, permet une appréciation plus nuancée et une *meilleure connaissance de l'élève*.

Le principal grief fait au nouveau système vient de professeurs de mathématiques qui lui reprochent son *manque de précision*.

7. — REMARQUES COMPLÉMENTAIRES FAITES PAR PLUSIEURS PROFESSEURS

Voici d'abord, par ordre de fréquence, trois remarques faites par un nombre assez élevé de professeurs.

as de
ponse

commentaire

VOUS SATISFAIT-IL PLUS QUE

seurs satisfaits du groupe des

1. Le nouveau système apporte un surcroît de travail administratif (conseils de classe, rapports, commentaires joints à la note, double bulletin).
2. L'échelle à cinq degrés n'est pas assez nuancée (à un même degré correspondent des scores trop différents).
3. La notation par chiffres influençait plus le rendement des élèves et était mieux comprise par les parents.

On relève encore :

- . Il faudrait mieux informer les parents.
- . Réformer le secondaire, oui ! Mais aussi le primaire !
- . Les enfants s'attribuent une trop grande liberté de travail.
- . Le passage de l'ancien système au nouveau aurait dû être plus progressif.
- . La notation varie selon les professeurs (personnalité) et selon les classes (niveau général de la classe).

Du côté positif, la plus grande justice, la plus grande humanité du nouveau système d'évaluation sont soulignées.

Parmi les remarques isolées, on note :

A. *Remarques négatives.*

- . Trop d'élèves par classe pour pratiquer efficacement le nouveau système.
- . Le nouveau système d'évaluation ne permet plus au professeur de sanctionner la paresse d'un élève par un retrait de points.
- . Le nouveau système permet au mauvais élève de dissimuler ses faiblesses (il se croit plus fort qu'il ne l'est en réalité).
- . Pourquoi n'adopte-t-on pas une méthode d'évaluation uniforme ?
- . Pourquoi évaluer par des exercices de contrôle quand on fait de l'évaluation continue par l'observation quotidienne ?
- . La suppression des examens, oui ! Le passage automatique, non !
- . Quels sont les critères pour attribuer un degré ? L'échelle n'est pas assez descriptive.
- . La matière est trop vaste et trop difficile.
- . Faire intervenir l'effort, l'activité et la conduite risque de fausser l'opinion qu'on peut se faire des aptitudes réelles de l'élève.

B. *Remarques positives.*

- . Le nouveau système est plus sé-
- . Il oblige l'enfant à travailler ré-
- . Il diminue les rivalités dans la
- . Il est plus nuancé.

G. — CC

La majorité des professeurs estiment que l'évaluation est un général, le nouveau système est sécurisant pour l'élève. Toutefois aux exigences d'une évaluation utilisable par l'élève.

Les principales difficultés s'expliquent par le manque de définition claire des finalités des outils, le manque d'outils adéquats. Qu'en est-il ? Nous n'avons pas obtenu les réponses suivantes :

- La note chiffrée doit-elle être utilisée ?
 - L'échelle ordinale convient-elle à l'évaluation de la performance ? Quels aménagements pour des échelles plus objectives et plus nuancées ?
- essai de réponse à ces deux de

B. *Remarques positives.*

- . Le nouveau système est plus sécurisant pour l'élève.
- . Il oblige l'enfant à travailler régulièrement.
- . Il diminue les rivalités dans la classe.
- . Il est plus nuancé.

G. — CONCLUSIONS

La majorité des professeurs auxquels nous nous sommes adressés estiment que l'évaluation est une de leurs fonctions essentielles. En général, le nouveau système est considéré comme plus humain et plus sécurisant pour l'élève. Toutefois, il ne paraît pas répondre entièrement aux exigences d'une évaluation rigoureuse, objective et directement utilisable par l'élève.

Les principales difficultés s'expliquent sans doute par l'absence d'une définition claire des finalités des nouveaux modes d'évaluation et par le manque d'outils adéquats. Que veut-on évaluer ? Avec quoi et avec qui ? Nous n'avons pas obtenu de réponses précises aux questions suivantes :

- La note chiffrée doit-elle être totalement abandonnée ?
- L'échelle ordinale convient-elle à l'évaluation de tous les types de performance ? Quels aménagements pourraient rendre les échelles plus objectives et plus nuancées ? La présente étude apporte un essai de réponse à ces deux dernières questions.

DEUXIÈME

Il est évident que les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

Les lois de la nature sont éternelles et immuables. Elles ne peuvent être ni créées ni détruites. Elles sont la base de toute existence et de toute action.

III. CONSTRUCTION D'ÉCHELLES D'ÉVALUATION DESCRIPTIVES

A. — DEUX GRADUS

Au-delà de toutes les nuances que l'on fait en se référant à une échelle généralement admise, toujours la même, qui consiste à classer simplement les objets considérés possédant plus ou moins d'une caractéristique.

Pour bien comprendre ces deux types d'évaluation, ce sera la route, à l'aide de système mesuré, l'usage du langage d'un groupe.

MESURE I

1. *Le critère est universel.*

Où que je me trouve dans le monde, quelle que soit ma culture ou mon caractère personnel, si je décide de mesurer la longueur en recourant au système métrique, je me référerai toujours au mètre-étalon (défini en fonction des radiations du krypton).

2. *Les intervalles de l'échelle sont égaux.*

Si un tronçon de route d'Amazonie mesure 200 km, alors que l'aller de Bruxelles-Liège n'en compte que 100, nous savons immédiatement que la route belge a la moitié de la longueur de la route amazonienne. Nous savons qu'en faisant le trajet Bruxelles-Liège et retour, nous parcourons une distance égale à la longueur du tronçon d'Amazonie.

X

CHAPITRE 1
LES ÉCHELLES D'ÉVALUATION

A. — DEUX GRANDS TYPES DE MESURES

Au-delà de toutes les nuances, il existe deux types de mesures : celles que l'on fait en se référant à une unité conventionnelle, parfois universellement admise, toujours la même pour ceux qui l'utilisent, et celles qui consistent à classer simplement, selon que l'individu ou la chose considérés possèdent plus ou moins tel attribut, telle propriété ou telle caractéristique.

Pour bien comprendre la différence capitale qui existe entre ces deux types d'évaluation, comparons la mesure de la longueur d'une route, à l'aide de système métrique (mesure I) à l'appréciation de l'élégance du langage d'un groupe d'élèves (mesure II).

MESURE I

MESURE II

1. *Le critère est universel.*

Où que je me trouve dans le monde, quelle que soit ma culture ou ma personnalité, si je décide de mesurer une longueur en recourant au système métrique, je me référerai toujours au même mètre-étalon (défini en fonction des radiations du krypton).

2. *Les intervalles de l'échelle sont tous égaux.*

Si un tronçon de route d'Amazonie mesure 200 km, alors que l'autoroute Bruxelles-Liège n'en compte que 100, nous savons immédiatement que la route belge a la moitié de la longueur de la route amazonienne. Nous savons aussi qu'en faisant le trajet Bruxelles-Liège et retour, nous parcourons une distance égale à la longueur du tronçon de route d'Amazonie.

1. *Il n'existe pas de critère universel.*

Même deux collègues de même formation, enseignant la langue maternelle au même niveau, dans la même école, se font une idée différente de l'élégance du langage. Tout au plus peuvent-ils se mettre d'accord sur certains indices, qu'ils tiennent pour signe de l'élégance. Qu'y-a-t-il de commun entre ce qu'un Bantou et un Latin considèrent comme langage élégant ?

2. *On ne sait rien sur la grandeur de chaque intervalle.*

Un professeur, jugeant de l'élégance du style, classe les compositions de ses élèves en cinq catégories : Très bien - Bien - Moyen - Faible - Insuffisant. Il est impossible de démontrer que la « quantité » d'élégance qui sépare « Moyen » de « Bien » est égale à la quantité entre « Bien » et « Très bien ». De plus, deux compositions affectées de la note « Bien » et écrites l'une à la suite

de l'autre ne deviennent pas une composition valant « Très bien ».

Si l'on considère maintenant deux professeurs de même formation, enseignant dans la même école, au même niveau, le « Très bien » accordé par l'un n'est pas égal à la même note accordée par l'autre. Que dire alors de « Très bien » accordés au Japon et en France !

3. Le zéro de l'échelle est absolu.

Une route de longueur zéro n'existe pas.

3. Il n'existe pas de zéro absolu.

Il est impossible de démontrer qu'une composition, si pauvre paraisse-t-elle, est *totalem*ent dépourvue d'élégance. Et ce qui paraît relâché aujourd'hui sera peut-être tenu pour une trouvaille demain.

Tout au plus peut-on dire que la composition classée dernière à un moment donné, occupe cette position *par rapport* aux autres, selon l'avis subjectif du professeur.

Alors que la mesure I se fait à l'aide d'une échelle de rapport, la mesure II recourt à une simple échelle ordinale. Et celle-ci n'indique jamais qu'un *rang*. Savoir que tel élève est le premier de sa classe en anglais, le plus grand par la taille, et le plus rapide en calcul mental n'apporte aucune précision réelle sur les connaissances en anglais, sur la grandeur ou sur l'aptitude numérique de l'élève en question : on peut être le plus grand des nains ou le plus petit des géants...

Dans la vie scolaire, de nombreuses évaluations ne sont que relatives. On classe les travaux, les performances, les comportements, les attitudes, du meilleur au moins bon, du plus grand au plus petit, du plus favorable au plus défavorable, etc. Les échelles d'évaluation à cinq degrés, adoptées dans l'enseignement secondaire rénové, sont de ce type.

Bref, l'échelle d'évaluation aide à ordonner des individus ou des objets par rapport à une qualité, une caractéristique, une propriété qu'ils possèdent à un degré plus ou moins élevé. C'est une *échelle ordinale*.

B. — ESPÈCES D'

1.— L'

La forme la plus élémentaire est arbitraire, représentant toute chose peut posséder une caractéristique par une simple marque.

EXEMPLE :

Connaissance de l'anglais p

2. —

Ex. 1 : Connaissance de l'an

Très bien
Bien
Satisfaisant
Faible
Insuffisant.

Ex. 2 : L'instituteur :

— tient toujours
× — tient parfois
— ne tient aucun

3. —

Connaissance de l'anglais

4. —

Connaissance de l'anglais

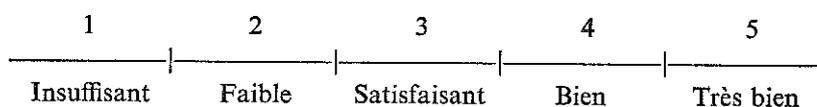
5. — L'ÉCHELLE CONCRÈTE

Dans ces échelles, les degrés sont concrétisés par des personnes ou des objets pris comme modèles.

EXEMPLE : 130 élèves ont été invités à dessiner le même objet. Pour évaluer, le professeur sélectionne le meilleur et le moins bon des dessins, puis un dessin « moyen ». Les dessins restants sont alors répartis en trois catégories, par comparaison avec les trois repères.

6. — L'ÉCHELLE COMBINÉE

Souvent, l'échelle graphique est combinée avec une échelle numérique, littérale ou verbale. Ci-dessous, toutes ces possibilités sont réunies.



7. — L'ÉCHELLE À CHOIX FORCÉ

Le juge doit choisir parmi une série de traits celui qui correspond le mieux à l'individu évalué.

EXEMPLE : Choisissez parmi les propositions suivantes celle qui décrit le mieux l'instituteur que vous venez d'observer :

- A. Tient toujours compte des remarques des élèves.
- B. Ne tient compte des remarques des élèves que si elles vont dans le sens de l'exposé.
- C. Poursuit son idée sans s'occuper des réactions de la classe.

Idéalement, les propositions d'une échelle à choix forcé doivent être rédigées de telle façon que les évaluateurs ne puissent pas déceler laquelle est la plus favorable par rapport à une option pédagogique ou, plus généralement, à un critère donné. On espère rendre ainsi toutes les propositions également attractives et réduire au maximum le jeu de la désirabilité sociale.

8. — L'ÉCHELLE (Self-a

Dans toutes les échelles trait à évaluer et les différe effectivement. Or s'il s'agit enseignant, on peut avoir de celui qui a construit l'éch ne comprenne même pas ce

Dans certains cas, il est l'évaluateur et de voir co (éventuellement lui-même)

F. P. Kilpatrick et H. Can schématiser de la façon suiv

1. Le sujet est invité à décri leur des enseignants. On que possible, jusqu'à ce
2. On procède de même p
3. On trace alors une éche à se situer lui-même ou

Les derniers types d'effort de précision du tra de la définition de critères formances scolaires.

C. — QUAND UTILIS

Selon la nature des ap tantôt un compte ou un c nale (note) s'imposent.

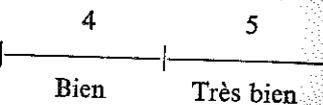
(¹) F. P. KILPATRICK et H. duals' unique reality worlds, REMMERS, in *Handbook of Re*

créés par des personnes ou

dessiner le même objet. Pour
ur et le moins bon des dessins,
estants sont alors répartis en
s trois repères.

MBINÉE

e avec une échelle numérique,
ces possibilités sont réunies.



IX FORCÉ

e traits celui qui correspond

ns suivantes celle qui décrit
observer :

des élèves.

es que si elles vont dans le

ctions de la classe.

e à choix forcé doivent être
puissent pas déceler laquelle
otion pédagogique ou, plus
ère rendre ainsi toutes les
e au maximum le jeu de la

8. — L'ÉCHELLE ANCRÉE PAR CELUI QUI ÉVALUE (Self-anchoring rating scale)

Dans toutes les échelles d'évaluation précédentes, la définition du trait à évaluer et les différents degrés sont imposés à celui qui évalue effectivement. Or s'il s'agit, par exemple, d'évaluer la qualité d'un enseignant, on peut avoir de cette qualité une conception très différente de celui qui a construit l'échelle; à la limite, il se peut que l'utilisateur ne comprenne même pas ce que le constructeur a voulu dire.

Dans certains cas, il est utile de connaître l'échelle de valeurs de l'évaluateur et de voir comment il situe l'objet de son observation (éventuellement lui-même) sur cette échelle. Dans cette perspective, F. P. Kilpatrick et H. Cantril ⁽¹⁾ propose une démarche qu'on peut schématiser de la façon suivante :

1. Le sujet est invité à décrire comment il conçoit, par exemple, le meilleur des enseignants. On pousse l'interview (enregistrée) aussi loin que possible, jusqu'à ce que le sujet n'ait plus rien à ajouter.
2. On procède de même pour définir le plus mauvais des enseignants.
3. On trace alors une échelle numérique sur laquelle le sujet est invité à se situer lui-même ou à situer un enseignant qu'il a observé.

* * *

Les derniers types d'échelles présentés témoignent d'un premier effort de précision du trait évalué. Nous reviendrons sur le problème de la définition de critères de mesure à propos de l'évaluation des performances scolaires.

C. — QUAND UTILISER LES ÉCHELLES D'ÉVALUATION ?

Selon la nature des apprentissages ou des performances à évaluer, tantôt un compte ou un décompte (score), tantôt une évaluation ordinaire (note) s'imposent.

⁽¹⁾ F. P. KILPATRICK et H. CANTRIL, Self-anchoring scaling. A measure of individuals' unique reality worlds, *J. of Indiv. Psychol.*, 1960, 16, 158-173. Cité par H. REMMERS, in *Handbook of Research on Teaching*, I, p. 364 sq.

En fait, il existe sans doute bien peu de cas où une performance scolaire ne puisse pas faire l'objet d'une évaluation ordinale. Par exemple, pour une dictée, il est fréquent d'assortir le score attribué par décompte, d'une note telle que « bien », « en progrès », etc.

La distinction entre *système de catégories* et *système de signes* (1) nous aidera à clarifier les notions.

- . Dans un système de *catégories*, on compte chaque occurrence d'un événement. Par exemple, on compte chaque intervention du professeur tendant à encourager un élève.
- . Dans un système de *signes*, l'événement n'est compté qu'une fois, mais fait l'objet d'une appréciation. Par exemple, on évalue dans quelle mesure un professeur a parlé clairement pendant sa leçon.

Le comptage, caractéristique du système de catégories, revêt un caractère nettement objectif; l'évaluation inhérente au système de signes est plus subjective. Cependant, à mesure que les critères d'une échelle d'évaluation se précisent un système de signes tend à se transformer en système de catégories.

Quand le recours aux échelles d'évaluation s'impose-t-il ?

L'élève à qui l'on pose une question peut être invité :

- soit à *construire* entièrement sa réponse,
 - . en toute liberté;
 - . ou en respectant certaines contraintes, par exemple de longueur.
- soit à *sélectionner* la réponse qu'il croit la meilleure parmi un choix proposé.

EXEMPLES :

A. *Sélection de réponses.*

1. Charles Quint est né en 1500 VRAI-FAUX
2. La capitale de la Hollande est :
 - Amsterdam — Rotterdam — La Haye.

(1) Voir à ce propos : G. DE LANDSHEERE, *Introduction à la recherche en éducation*, Paris, A. Collin; Liège, G. Thone, 1976, 4^e éd.

3. L'articulation radio-cubito-carpale et la rotation limitée PARCE que le ligament annulaire.

- A. La constatation et la relation de cause à effet.
- B. La constatation et la relation de cause à effet.
- C. La constatation est vraie.
- D. La constatation est fautive.
- E. La constatation et la relation de cause à effet.

4. Exemple où plusieurs r...
Cinq conscrits mesurent

1. La moyenne de tailles est de 1,70 m.
 2. L'écart type est proche de 0,05 m.
 3. L'écart type est proche de 0,10 m.
 4. L'échantillon a de bonnes tailles des individus en moyenne.
- A. 1, 2, 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.

B. *Construction de réponses.*

1. Réponse entièrement ou partiellement correcte.
Ecoutez le *Prélude à l'opéra* et notez vos impressions.
2. Réponse soumise à des contraintes.
 - . Simple réponse de forme fixe.
— Exemple : la forme d'un triangle.
 - . Rechercher les jugements.
— Dans le même esprit que les exercices d'analyse, les exercices d'analyse.
— découvrir des relations.
— décider si des conclusions sont vraies.
— découvrir les cas particuliers.
— interpréter des cas particuliers.

ÉVALUATION DES RÉPONSES.

Dans le système de sélection, les questions sont construites purement factuelles et la correction est aisée; qu...

cas où une performance
ation ordinale. Par exemple,
ore attribué par décompte,
, etc.

s et système de signes (1)

chaque occurrence d'un
intervention du professeur

n'est compté qu'une fois,
exemple, on évalue dans
ement pendant sa leçon.

de catégories, revêt un
rente au système de signes
les critères d'une échelle
s tend à se transformer en

n s'impose-t-il ?

être invité :

par exemple de longueur.

meilleure parmi un choix

-FAUX

La Haye.

on à la recherche en éducation,

3. L'articulation radio-cubitale supérieure permet des mouvements de rotation limités PARCE QUE la tête du radius est entourée par le ligament annulaire.

- A. La constatation et la raison proposée sont toutes les deux vraies et il existe une relation de cause à effet entre les deux.
- B. La constatation et la raison proposée sont toutes les deux vraies et il n'y a pas de relation de cause à effet entre elles.
- C. La constatation est vraie, mais la raison proposée est fausse.
- D. La constatation est fausse, mais la raison proposée est un fait ou un principe accepté.
- E. La constatation et la raison proposée sont toutes les deux fausses.

4. Exemple où plusieurs réponses sont correctes :

Cinq conscrits mesurent : 1,65 m - 1,67 m - 1,69 m - 1,63 m et 1,61 m.

- 1. La moyenne de 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, 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.

B. Construction de réponses.

1. Réponse entièrement ouverte :

Ecoutez le *Prélude à l'après-midi d'un faune*, de Debussy et exprimez vos impressions.

2. Réponse soumise à des contraintes :

- Simple réponse de fait :
— Exemple : la formule chimique du sel de cuisine s'écrit...
- Rechercher les jugements de valeur dans un texte.
- Dans le même esprit que l'exemple précédent, et de façon plus générale, les exercices d'analyse et d'interprétation peuvent consister à :
— découvrir des relations;
— décider si des conclusions sont valides ou non;
— découvrir les cas d'application appropriée de certains principes;
— interpréter des cartes, des graphiques,...

ÉVALUATION DES RÉPONSES.

Dans le système de sélection de réponses ou lorsqu'il s'agit de réponses construites purement factuelles ou soumises à d'étroites contraintes, la correction est aisée; quand il s'agit de comptage d'éléments entière-

ment définis, la correction peut être automatisée. Ainsi, à la rapidité générale de la réponse s'ajoute la rapidité de la correction.

Il en va autrement lorsque l'élève répond en toute liberté. A la limite, il crée ainsi une œuvre propre, unique, dont l'évaluation échappe aux simples comptages.

C'est ici que l'échelle d'évaluation va jouer pleinement son rôle. Nous l'avons déjà vu, *l'échelle d'évaluation concerne surtout les performances complexes.*

NIVEAU COGNITIF.

Les exemples précédents montrent que les questions à sélection de réponses (choix multiple) ne se cantonnent pas nécessairement à la simple mémoire des faits et peuvent mettre en jeu des processus intellectuels supérieurs.

Il serait néanmoins aberrant de donner à tous les exercices scolaires la forme de questions à choix multiple. Il importe que l'élève sache aussi s'exprimer constructivement, voire créativement ⁽¹⁾. Dans ce type de réponses, on ne s'intéresse pas seulement à des capacités particulières, mais aussi, sinon surtout, à la façon dont l'élève les conjugue, les articule, pour produire un tout intégré. La complexité même de ces réponses explique la difficulté de leur évaluation. Et cette difficulté est accrue du fait que les objectifs tendent à s'exprimer d'autant plus vaguement que la performance qu'ils impliquent est complexe (exemple : développer l'esprit critique).

Quoi qu'il en soit, *les questions à réponses librement construites* ⁽²⁾ *doivent donc continuer à jouer à plein leur rôle éducatif. Les échelles d'évaluation sont les plus adéquates pour évaluer ces productions libres.*

⁽¹⁾ La prise de position de J. P. GUILFORD est, dans ce contexte, fort éclairante : « L'usage exclusif d'épreuves à réponses standardisées n'encourage pas l'effort de pensée productive. En réalité, il le décourage en renforçant d'autres espèces de performances appelées par les réponses standardisées (...). Il est alarmant d'imaginer les conséquences que pourrait avoir l'usage exclusif d'épreuves à réponses standardisées sur le caractère intellectuel d'une nation. »

Cf. J. P. GUILFORD, *The Nature of Human Intelligence*, New York, McGraw-Hill, 1967, p. 476.

⁽²⁾ Ce que les Anglo-Saxons appellent les *essay tests*, expression qui n'a pas de traduction en français.

D. — PHÉNOMÈNES ET LA VALIDITÉ

Nous l'avons déjà vu, si les juges ne sont pas suffisamment précisés, les jugements sont instables. Dans la littérature expérimentale, on a étudié des phénomènes plus particuliers.

1. —

Le juge se laisse influencer par une certaine élégance de langage, par ses préjugés sociaux.

EXEMPLE : On tend à surestimer

2. — L'EFFET

Savoir qu'un élève a réussi une « distinction » entraîne une surévaluation.

3. — L'EFFET

C'est un cas de contamination. Beaucoup d'élèves ont vu se réaliser une performance obtenue lors d'un premier examen.

Quand on invite à évaluer une production à l'aide d'une batterie d'échelles, on observe parfois une contamination d'un même côté.

4. — L'EXCÈS DE

5. — LA

Parce que les élèves moyens sont sans doute aussi pour éviter un mauvais coup de juges situent leur

D. — PHÉNOMÈNES MENAÇANT L'OBJECTIVITÉ ET LA VALIDITÉ DES ÉVALUATIONS

Nous l'avons déjà vu, si les objectifs de l'évaluation ne sont pas suffisamment précisés, les juges adoptent des critères différents et peu stables. Dans la littérature expérimentale, on trouve la description de phénomènes plus particuliers.

1. — L'EFFET DE HALO

Le juge se laisse influencer par l'aspect extérieur de l'élève, par une certaine élégance de langage, par la présentation matérielle des travaux, par ses préjugés sociaux.

EXEMPLE : On tend à surestimer l'intelligence d'un élève qui a l'œil vif.

2. — L'EFFET DE CONTAMINATION

Savoir qu'un élève a réussi ses examens précédents « avec grande distinction » entraîne une surévaluation chez certains éducateurs.

3. — L'EFFET DE STÉRÉOTYPIE

C'est un cas de contamination particulier. En dépit de leurs efforts, beaucoup d'élèves ont vu se répéter toute une année la note qu'ils avaient obtenue lors d'un premier exercice.

Quand on invite à évaluer des caractéristiques ou des traits différents à l'aide d'une batterie d'échelles présentées les unes en dessous des autres, on observe parfois une tendance à placer la majorité des appréciations d'un même côté.

4. — L'EXCÈS DE SÉVÉRITÉ OU DE GÉNÉROSITÉ

5. — LA TENDANCE CENTRALE

Parce que les élèves moyens sont en général les plus nombreux et sans doute aussi pour éviter de prendre trop de responsabilités, beaucoup de juges situent leur évaluation autour du point « moyen ».

(Dans ce cas, les professeurs qui osent utiliser les degrés extrêmes de l'échelle pèsent anormalement dans les décisions de réussite ou d'échec.)

6. — L'EFFET DE CONTRASTE

On tend, par exemple, à sous-évaluer un élève interrogé immédiatement après un condisciple très brillant.

7. — LA DISTRIBUTION FORCÉE

A la suite de regrettables confusions entre la répartition normale des aptitudes et les résultats d'un enseignement réussi, entre la sélection et l'éducation fondamentale, d'aucuns ont souhaité que les résultats scolaires se distribuent selon la courbe de Gauss. On a, par exemple, préconisé que, sur 20 élèves, 1 soit classé très bon, 5 bons, 8 moyens, 5 médiocres et 1 faible. On préjuge ainsi des résultats de l'enseignement qui, idéalement, devrait conduire tous les élèves à un très bon résultat (en nivelant par le haut !).

Les recherches illustrant la contamination des notes par des *facteurs externes* sont surtout nombreuses dans le domaine de la composition écrite. En voici quelques exemples.

John C. Marshall (1) a montré l'influence de certaines variables formelles sur la note attribuée par un examinateur. Une copie de sciences sociales jugée de valeur moyenne par des experts est sélectionnée. Treize formes différentes de cette copie sont préparées sans jamais modifier le contenu, mais en introduisant des fautes d'orthographe ou de grammaire :

- 3 formes comportent 6, 12 ou 18 fautes d'orthographe;
- 3 formes comportent 6, 12 ou 18 erreurs de grammaire;
- 3 formes comportent 6, 12 ou 18 erreurs de ponctuation;
- 3 formes combinées comportent 2, 4 ou 6 erreurs de chaque type;
- 1 forme de contrôle ne comporte aucune erreur d'orthographe, de ponctuation ou de grammaire.

(1) John C. MARSHALL, Composition error and essay examination grades re-examined, *Journal of Educational Measurement*, 1972, vol. 5.

Sept cents professeurs sont choisis en fonction de leur nom et de leur spécialité. Chaque treize formes, à l'aide d'une grille de correction, du contenu de la réponse.

Cette expérience permet de

1. Bien qu'il leur soit instamment demandé de ne pas modifier de la réponse, les professeurs ont introduit des fautes formelles de la copie : on observe un nombre de fautes d'orthographe et de composition. Les erreurs plus affecter la note que le nombre d'erreurs n'est pas. 18 erreurs sont sous-évaluées, 6 erreurs ou moins.
2. Les maîtres qui ont une réputation plus sévères.
3. Les professeurs les plus sociaux sont les plus sévères.

Dans une expérience de ce genre, on a fait varier la présentation de la grille de correction, d'autres conditions suivantes :

1. La qualité de l'écriture. Les juges sont plus sévères pour une mauvaise qualité.
2. On observe un effet de sévérité pour les mal écrites lorsqu'elles sont comparées à des copies dont l'écriture est bonne.
3. Les fautes d'orthographe.

(1) C. CHASE, The impact of handwriting on the grade of Educational Measurement, 1972, vol. 5.

(2) Contrairement à l'expérience de Marshall, la qualité d'orthographe n'a pas varié systématiquement.

Sept cents professeurs sont appelés à juger ces compositions; ils sont choisis en fonction de leur nombre d'années d'expérience d'enseignement et de leur spécialité. Chaque professeur est invité à évaluer une des treize formes, à l'aide d'une échelle à neuf degrés, en ne se souciant que du contenu de la réponse.

Cette expérience permet les observations suivantes :

1. Bien qu'il leur soit instamment demandé de n'évaluer que le contenu de la réponse, les professeurs sont influencés par les caractéristiques formelles de la copie : on observe une relation inverse entre le nombre de fautes d'orthographe ou de grammaire et la note accordée à la composition. Les erreurs d'orthographe et de grammaire semblent plus affecter la note que les erreurs de ponctuation. L'influence du nombre d'erreurs n'est pas uniforme. Les copies contenant 12 ou 18 erreurs sont sous-évaluées par rapport à celles qui comportent 6 erreurs ou moins.
2. Les maîtres qui ont une plus longue expérience pédagogique sont les plus sévères.
3. Les professeurs les plus spécialisés dans le domaine des sciences sociales sont les plus sévères.

Dans une expérience du même genre, C. Chase ⁽¹⁾ a aussi montré combien la présentation d'une copie influence les juges. Cette fois, on a fait varier l'écriture et l'orthographe. Certains juges utilisent une grille de correction, d'autres n'en disposent pas. On retiendra les conclusions suivantes :

1. La qualité de l'écriture a une influence significative sur la note finale. Les juges sont plus généreux lorsque l'écriture est de très bonne qualité.
2. On observe un effet de halo négatif (sous-évaluation) pour les copies mal écrites lorsqu'elles sont évaluées immédiatement après des copies dont l'écriture est bonne, et inversement.
3. Les fautes d'orthographe n'influencent pas la note finale ⁽²⁾.

⁽¹⁾ C. CHASE, The impact of some obvious variables on essay test scores, *Journal of Educational Measurement*, n° 4, 1968.

⁽²⁾ Contrairement à l'expérience décrite précédemment, le nombre de fautes d'orthographe n'a pas varié systématiquement dans cette expérience.

4. Les juges qui utilisent la grille de correction attribuent plus de points que ceux qui ne l'utilisent pas.

E. — NÉCESSITÉ DE DÉFINITIONS OPÉRATIONNELLES

Même si des techniques particulières ⁽¹⁾ permettent de pallier l'effet des facteurs externes, pourquoi la fidélité des évaluations d'un même juge ou la concordance entre plusieurs juges pour une même performance reste-t-elle souvent aussi basse lorsqu'on utilise les échelles d'appréciation classiques ?

D'abord parce que beaucoup d'entre elles comptent un trop grand nombre de degrés. Il n'existe pas de règle formelle, mais on observe généralement qu'un juge ne réussit pas à classer, avec quelque sûreté, en plus de onze catégories. Les échelles à cinq degrés se révèlent sûres, mais, comme le confirme notre enquête préliminaire, elles donnent souvent aux juges l'impression de ne pas pouvoir assez nuancer leurs appréciations. L'expérience montre qu'il reste cependant préférable de commencer par juger selon cinq catégories puis d'ajouter, au besoin, une nuance vers le haut ou vers le bas. Par exemple, dans la tradition académique belge où l'on classe de la plus grande distinction à l'échec, en passant par la grande distinction, la distinction et la satisfaction, on entend souvent parler, en délibération du jury, d'une « belle distinction » ou, au contraire, d'une « petite distinction ».

Certains auteurs ⁽²⁾ considèrent qu'il est dangereux d'adopter des échelles comportant un nombre impair de degrés parce que les juges ont tendance à porter leurs évaluations sur l'échelon médian (effet de tendance centrale). W. McColly et R. Remstad ont expérimenté l'efficacité relative d'échelles non descriptives à quatre ou six niveaux pour l'évaluation de compositions. Ces deux instruments ne se différencient pas du point de vue de la fidélité et de la sensibilité. L'échelle à quatre

⁽¹⁾ Voir 1^{re} partie, p. 25.

⁽²⁾ W. MCCOLLY et R. REMSTAD, Composition rating scale for general merit : an experimental evaluation, *Journal of Educational Research*, vol. 59, n° 2, octobre 1965.

niveaux présente seulement l'a rapide.

En fait, ces controverses sur que si l'on utilise des échelles n du manque de fidélité dans l'us réside ailleurs : elle se trouve d Inviter à ordonner les élèves en 3, 4, 5, ou Très bien, bien, etc., d'interpréter ces symboles ou moment. Et le fait que certains aisément à se mettre d'accord faire oublier les désaccords gra

Les solutions sûres à ce prob qu'une : *s'accorder sur une déff luation et sur chacun des deg critères.*

Sans cet effort d'opération différents et les facteurs intern leur influence perturbatrice.

⁽⁴⁾ « La définition opérationnelle des procédures qui permettent de reconnaître parmi d'autres. » H.

on attribuent plus de points

OPÉRATIONNELLES

permettent de pallier l'effet
des évaluations d'un même
pour une même perfor-
qu'on utilise les échelles

s comptent un trop grand
formelle, mais on observe
passer, avec quelque sûreté,
à cinq degrés se révèlent
enquête préliminaire, elles
pas pouvoir assez nuancer
il reste cependant préfé-
catégories puis d'ajouter,
bas. Par exemple, dans la
la plus grande distinction
1, la distinction et la satis-
tion du jury, d'une « belle
distinction ».

dangereux d'adopter des
degrés parce que les juges
l'échelon médian (effet de
ad ont expérimenté l'effi-
quatre ou six niveaux pour
niveaux ne se différencient
sibilité. L'échelle à quatre

rating scale for general merit :
research, vol. 59, n° 2, octobre

niveaux présente seulement l'avantage d'être d'une application plus rapide.

En fait, ces controverses sur le nombre d'échelons n'ont de sens que si l'on utilise des échelles non descriptives. Car la cause essentielle du manque de fidélité dans l'usage des échelles, quelles qu'elles soient, réside ailleurs : elle se trouve dans le *vague des critères de classement*. Inviter à ordonner les élèves en cinq catégories, A, B, C, D, E ou 1, 2, 3, 4, 5, ou Très bien, bien, etc., c'est laisser à chaque évaluateur le soin d'interpréter ces symboles ou ces mentions selon son inspiration du moment. Et le fait que certains enseignants chevronnés réussissent assez aisément à se mettre d'accord sur certaines appréciations ne doit pas faire oublier les désaccords graves qui se produisent aussi entre eux.

Les solutions sûres à ce problème ne sont pas nombreuses. Il n'y en a qu'une : *s'accorder sur une définition opérationnelle (1) des critères d'évaluation et sur chacun des degrés de l'échelle relative à chacun de ces critères*.

Sans cet effort d'opérationnalisation, les juges adoptent des critères différents et les facteurs internes et externes, définis ci-dessus, exercent leur influence perturbatrice.

(1) « La définition opérationnelle d'un comportement, d'un facteur, etc. est l'énoncé des procédures qui permettent de le mesurer, de le produire ou simplement de le reconnaître parmi d'autres. » H. PIERON, *Vocabulaire de la psychologie*, Paris, P.U.F.

CHAPITRE 2

ÉTAPES ET CONDITIONS DE LA RECHERCHE

A. — ÉTAPES

Nous avons déjà insisté sur la nécessité d'associer les enseignants à la construction des instruments d'évaluation. Réunis en commissions, inspecteurs et professeurs jouent le rôle d'experts à un double point de vue. D'une part, eux seuls possèdent une connaissance suffisamment approfondie de la matière pour guider le choix des critères d'évaluation et en saisir toutes les implications. D'autre part, leur expérience permet de fixer les activités et les exigences les mieux adaptées à un niveau pédagogique donné. Dans la présente recherche, des commissions ont fonctionné à Liège et dans la province du Luxembourg.

L'objectif était la mise au point d'une méthodologie de construction des échelles descriptives et la reconnaissance de l'efficacité et des limites de ces instruments. Seule, une expérience menée au premier cycle de l'enseignement secondaire rénové est relatée en détail; les commentaires revêtiront cependant un caractère plus général.

Les étapes de la mise au point des échelles pour la première année de l'enseignement rénové ont été les suivantes :

- Rédaction et prétest;
- Expérimentation sur grand nombre;
- Interviews destinées à vérifier les hypothèses suggérées par les résultats de l'expérimentation;
- Remaniement des instruments et nouvelle expérimentation.

B. — CONDITIONS

1. — DISCIPLINES CONCERNÉES

L'échelle descriptive étant un instrument adapté à l'évaluation de performances complexes, il s'agissait donc de choisir, dans chacune des branches, un exercice qui se prête à une large palette d'évaluations et qui soit régulièrement utilisé dans les classes. En langue maternelle,

la composition française répond
matique, par contre, il s'est avé
parmi ceux que l'on pratique
avons donc dû nous résoudre
autre discipline du type scientifi
un rapport d'expérience prop
scientifiques expérimentaux.

2. — LE

En langue maternelle, un tex
rédaction traditionnelle. On p
policier : un résumé du déb
doit s'insérer la composition

La préférence a été accord
contraintes imposées par le
fournissent des points de r
laissant une place suffisante
présenté aux élèves est inclu

En sciences, le professeur
relative à la dureté de l'eau
sur les produits utilisés. On
phénomènes observés, de l
prêter et d'inventer une mé
périence de base est décrite

A titre indicatif, nous pr
le second cycle de l'enseig
nelle, les élèves ont été invi
« Propos sur l'éducation »
élèves d'observer une expé
halogènes), de l'interprét

(¹) Des recherches bibliogra
inspecteurs et enseignants nous
solicitant chez l'élève des per
inconvenient d'être totalemen

(²) Proposé par G. LEROY.

A RECHERCHE

associer les enseignants à
Réunis en commissions,
erts à un double point de
onnaissance suffisamment
des critères d'évaluation
rt, leur expérience permet
ux adaptées à un niveau
he, des commissions ont
embourg.

odologie de construction
e l'efficacité et des limites
e au premier cycle de l'en-
détail; les commentaires

pour la première année

s suggérées par les résul-

xpérimentation.

NÉES

dapté à l'évaluation de
e choisir, dans chacune
rge palette d'évaluations
s. En langue maternelle,

la composition française répond parfaitement à ces exigences. En mathématique, par contre, il s'est avéré impossible de trouver l'exercice voulu parmi ceux que l'on pratique couramment dans les classes (1). Nous avons donc dû nous résoudre à chercher l'exercice adéquat dans une autre discipline du type scientifique. Le choix s'est finalement porté sur un rapport d'expérience proposé dans le cadre du cours de travaux scientifiques expérimentaux.

2. — LES EXERCICES CHOISIS

En langue maternelle, un texte mutilé en son milieu a été préféré à la rédaction traditionnelle. On présente à l'élève les éléments d'un roman policier : un résumé du début, puis deux paragraphes entre lesquels doit s'insérer la composition de l'élève.

La préférence a été accordée à ce genre d'exercice (2) parce que les contraintes imposées par le résumé et les paragraphes d'encadrement fournissent des points de repère objectifs pour l'évaluation, tout en laissant une place suffisante à l'apport spontané de l'élève. Le texte présenté aux élèves est inclus dans le dossier figurant en annexe.

En sciences, le professeur réalise, devant les élèves, une expérience relative à la dureté de l'eau, en fournissant seulement des informations sur les produits utilisés. On demande à l'élève de décrire par écrit les phénomènes observés, de les représenter par un schéma, de les interpréter et d'inventer une méthode de mesure de la dureté de l'eau. L'expérience de base est décrite dans le dossier figurant en annexe.

A titre indicatif, nous présentons, en outre, les exercices choisis pour le second cycle de l'enseignement secondaire rénové. En langue maternelle, les élèves ont été invités à résumer et à critiquer un texte d'Alain : « Propos sur l'éducation ». En sciences, on a proposé de nouveau aux élèves d'observer une expérience (déplacement d'halogènes par d'autres halogènes), de l'interpréter et d'imaginer une méthode d'analyse.

(1) Des recherches bibliographiques et de nombreuses séances de travail réunissant inspecteurs et enseignants nous ont permis de construire un exercice de mathématique sollicitant chez l'élève des performances complexes. Il présentait toutefois le grave inconvénient d'être totalement étranger au contexte scolaire habituel.

(2) Proposé par G. LEROY.

AVERTISSEMENT

Avant d'aborder le chapitre suivant, relatif à la méthodologie de la construction des échelles descriptives, le lecteur prendra utilement connaissance d'un des dossiers figurant en annexe ⁽¹⁾. En effet, l'exposé méthodologique est illustré d'extraits de ces dossiers et ces exemples ne sont vraiment parlants que dans la mesure où l'on en connaît le contexte.

Convaincus de l'efficacité de l'exercice personnel, nous avons joint une réponse d'élève aux dossiers du premier cycle. Le lecteur trouve ainsi immédiatement l'occasion de s'exercer.

Pour beaucoup, la démarche proposée rompt des habitudes professionnelles et paraît, au premier abord, assez lourde. L'expérience montre qu'après quelques corrections, de nouvelles habitudes se créent et allègent considérablement le travail.

⁽¹⁾ Chacun de ces dossiers fait l'objet d'une publication séparée.

MÉTHODOLOGIE DE LA D'ÉVALUATION

A. — CONSTRUCTION

Les échelles d'évaluation ou l'objet de nombreuses recherches essentiellement sur l'utilisation de la méthodologie de la construction ou, au moins, les recommandations à proposer doivent donc être suggérées de l'emploi ou de l'expérience.

La méthodologie de construction de l'aboutissement de nombreux échelons prétention d'être achevée. En

- La sélection et la définition
- La sélection et la définition d'un critère.
- La détermination des niveaux
- La description des échelons
- La présentation de l'échelle

Chacune de ces étapes va donc être reprise aux dossiers présentés dans l'autre en sciences (premier cycle)

I. —

Sélection et définition

a) UN PETIT NOMBRE DE CRITÈRES

On pourrait croire que l'on est obligé à prendre de nombreux

CHAPITRE 3

MÉTHODOLOGIE DE LA CONSTRUCTION D'ÉCHELLES D'ÉVALUATION DESCRIPTIVES

A. — CONSTRUCTION INITIALE DE L'INSTRUMENT

Les échelles d'évaluation ou d'appréciation (*rating scales*) ont fait l'objet de nombreuses recherches expérimentales. Celles-ci portent essentiellement sur l'utilisation. Par contre, la littérature relative à la méthodologie de la construction des échelles est fort pauvre. Les règles ou, au moins, les recommandations ou les conseils que l'on voudrait proposer doivent donc être suggérés par les difficultés rencontrées lors de l'emploi ou de l'expérimentation d'échelles existantes.

La méthodologie de construction, esquissée par la suite, constitue l'aboutissement de nombreux essais et tâtonnements. Elle n'a pas la prétention d'être achevée. En gros, on distingue cinq étapes :

- La sélection et la définition des critères de l'évaluation.
- La sélection et la définition des aspects qui interviennent dans chaque critère.
- La détermination des niveaux d'exigence pour chaque aspect.
- La description des échelons.
- La présentation de l'échelle et la rédaction des consignes d'utilisation.

Chacune de ces étapes va donc être décrite et illustrée par des exemples repris aux dossiers présentés en annexe : l'un en langue maternelle et l'autre en sciences (premier cycle de l'enseignement rénové).

I. — PREMIÈRE ÉTAPE

Sélection et définition des critères d'évaluation

a) UN PETIT NOMBRE DE CRITÈRES.

On pourrait croire que l'évaluation d'une performance complexe oblige à prendre de nombreux critères en considération. Or il est presque

toujours peu indiqué de procéder ainsi, car la précision espérée et les distinctions subtiles qui l'accompagnent sont souvent illusoire

De nombreuses expériences montrent que des correcteurs non avertis croient souvent tenir compte d'un grand nombre de critères, alors qu'en réalité, et sans s'en rendre compte, ils envisagent plusieurs fois les mêmes, mais en les baptisant différemment. La recherche de C. Remondino (1) illustre bien ce phénomène. Remondino a invité vingt professeurs de langue maternelle à dresser la liste des caractéristiques dont ils tiennent compte dans la correction des compositions. Une analyse factorielle montre que les dix-sept caractéristiques citées se réduisent en réalité à quatre : présentation, usage de la langue, contenu et organisation des idées, originalité.

Dans toute classification, dont l'évaluation est une des formes, les catégories choisies doivent répondre à l'exigence fondamentale de l'exclusion mutuelle (une même observation ne doit pas pouvoir entrer dans deux catégories à la fois). Respecter cette exigence est difficile et les transgressions ne sont pas aisément décelables. Ainsi, dans une échelle proposée par l'Educational Testing Service (2), pour la notation de compositions de langue maternelle, des recouvrements apparaissent dans la définition de deux catégories.

La description du niveau moyen de l'échelle « Idées ou contenu » pénalise la présence d'idées conventionnelles : « Les idées sont appropriées, mais conventionnelles et peu nombreuses... » Par ailleurs, une échelle est consacrée à l'évaluation de la créativité. Son niveau moyen se définit comme suit : L'étudiant (...) traite de choses à peu près comme tout le monde... (p. 150).

Vouloir tout évaluer dans les exercices faits par les élèves serait à la fois peu économique et peu rentable. Peu économique parce que le temps consacré par le professeur à la correction des travaux serait excessif; peu rentable parce qu'un élève ne peut pas porter simultanément ses efforts sur la correction d'un grand nombre de lacunes.

Par ailleurs, plusieurs études montrent que des juges évaluant une

(1) Cité par G. DE LANDSHEERE, *Evaluation continue et examens. Précis de didactologie*, Bruxelles, Labor; Paris, Nathan, 4^e édition, 1976, p. 146.

(2) Cf. G. DE LANDSHEERE, *o.c.*, p. 148.

L'Educational Testing Service est actuellement la plus grande entreprise de construction d'instruments d'évaluation à usage scolaire.

production d'élèves sans en avoir retenu presque toujours les caractéristiques risquent ainsi de n'être

Dès lors, comment choisir les

La méthode que nous avons proposée aux professeurs enseignant la discipline de la langue maternelle, réunis en plusieurs groupes, a permis de débattre et de discuter assez rapidement l'essentiel de la question.

La méthode proposée par H. Abeles invite des professeurs à évaluer une performance d'élève qui est obtenue dans une situation. Les résultats obtenus sont alors soumis à une analyse. Souvent une longue liste de critères et de juges sont invités à apprécier la performance à cinq niveaux. Les résultats sont analysés afin de découvrir les principaux

Evidemment, même réduites à cinq caractéristiques d'un travail, les caractéristiques d'un travail sont souvent plus nombreuses que son contenu. On peut aussi évaluer un élève par rapport à d'autres. On peut aussi évaluer à la faculté de synthétiser des idées disparates, puisés dans la mémoire.

Une des façons de marquer un élève consiste à lui consacrer un point de vue qu'aux autres. Dans le dossier de rapport scientifique demandé par l'enseignant, on a consacré le même

(1) E.T.S., *Definition of ratings for the analysis of essays by computer*, 1318, 1968.

(2) H. ABELES, A faced-factorial design for the measurement of complex behaviors, *Journal of Educational Psychology*, 1968.

la précision espérée et les
souvent illusoire

es correcteurs non avertis
re de critères, alors qu'en
t plusieurs fois les mêmes,
che de C. Remondino (1)
vité vingt professeurs de
ristiques dont ils tiennent
Une analyse factorielle
se réduisent en réalité à
tenu et organisation des

est une des formes, les
nce fondamentale de l'ex-
t pas pouvoir entrer dans
igence est difficile et les
. Ainsi, dans une échelle
pour la notation de com-
ents apparaissent dans la

dées ou contenu » pénalise
ont appropriées, mais con-
ne échelle est consacrée à
définit comme suit : L'étu-
le monde... (p. 150).

par les élèves serait à la
conomique parce que le
ction des travaux serait
ut pas porter simultanément
nombre de lacunes.

des juges évaluant une

et examens. Précis de doci-
76, p. 146.

grande entreprise de construc-

production d'élèves sans en avoir préalablement spécifié les objectifs,
retiennent presque toujours les mêmes critères (1). Certaines caracté-
ristiques risquent ainsi de n'être jamais prises en considération.

Dès lors, comment choisir les critères pertinents pour évaluer un travail ?

La méthode que nous avons retenue consiste à consulter des profes-
seurs enseignant la discipline à un niveau pédagogique donné. Ces
professeurs, réunis en plusieurs groupes travaillent simultanément et
débattent des critères à retenir. Cette façon de faire permet de repérer
assez rapidement l'essentiel des caractéristiques à prendre en consi-
dération.

La méthode proposée par Abeles (2) systématise cette démarche.
Abeles invite des professeurs à décrire, de la façon la plus précise possible,
une performance d'élève qui leur est proposée. Les descriptions ainsi
obtenues sont alors soumises à une analyse du contenu qui fournit
souvent une longue liste de critères. Pour chacun de ces critères, les
juges sont invités à apprécier des travaux d'élèves en utilisant une échelle
à cinq niveaux. Les résultats sont ensuite traités par analyse factorielle,
afin de découvrir les principaux critères de l'évaluation.

Evidemment, même réduites à leur plus simple expression, toutes
les caractéristiques d'un travail d'élève ne revêtent pas la même impor-
tance. Par exemple, la présentation d'une composition importe moins
que son contenu. On peut aussi désirer privilégier un processus intellec-
tuel par rapport à d'autres. On accordera, par exemple, plus d'attention
à la faculté de synthétiser des idées qu'à la simple accumulation de faits
disparates, puisés dans la mémoire.

Une des façons de marquer l'importance d'un processus particulier
consiste à lui consacrer un plus grand nombre d'échelles d'évaluation
qu'aux autres. Dans le dossier annexé, on verra que, pour apprécier le
rapport scientifique demandé au premier cycle de l'enseignement secon-
daire, on a consacré le même nombre d'échelles aux processus de rappel

(1) E.T.S., *Definition of ratings on the ETS Compositions Scale*, cité par E. PAGE,
The Analysis of Essays by Computer, Washington, U.S. Office of H.E.W., Project 6-
1318, 1968.

(2) H. ABELES, A faced-factorial approach to the construction of rating scales to
measure complex behaviors, *Journal of Educational Research*, summer 1973.

(relation de l'expérience par le texte et le schéma) qu'aux processus supérieurs (interprétation et invention). Par contre, au deuxième cycle, une place plus grande est réservée aux processus supérieurs; deux échelles leur sont consacrées, alors que la relation de l'expérience n'est considérée qu'une fois.

b) COMMENT DÉFINIR LES CRITÈRES ?

Dans la méthode que nous avons suivie, la première définition que l'on donne d'un critère reste assez générale. Elle se précise progressivement par le choix des aspects et par la détermination des niveaux d'exigence. Examinons les exemples suivants :

— LANGUE MATERNELLE : *définition des critères de l'évaluation.*

- a) *Pertinence sémantique* : le travail est pertinent au point de vue sémantique si l'élève développe des idées appropriées au contexte c'est-à-dire compatibles avec le résumé et les paragraphes 1 et 3.
- b) *Pertinence syntaxique* : respect des contraintes syntaxiques imposées par les paragraphes 1 et 3.
- c) *Vocabulaire* : ensemble des mots utilisés par l'élève. On ne tient pas compte de l'orthographe.
- d) *Structures grammaticales* : Le critère est ici la syntaxe du paragraphe rédigé par l'élève, indépendamment des contraintes imposées par les paragraphes 1 et 3 (le respect de ces contraintes a déjà été évalué en b).
- e) *Organisation des idées dans le paragraphe 2* : on évalue la structuration du travail, indépendamment du respect des contraintes imposées par les paragraphes 1 et 3 (le respect de ces contraintes a déjà été évalué en a).

— SCIENCES : *définition des critères de l'évaluation.*

- a) *Description écrite de l'expérience* : traduction, par le texte, des phénomènes observés.
- b) *Description de l'expérience par un ou plusieurs croquis* : on évalue les schémas présentés par l'élève et les annotations qui les complètent.
- c) *Interprétation de l'expérience* : on évalue les commentaires des élèves qui, au-delà de la description des étapes de l'expérience, mettent en évidence les relations entre les phénomènes observés pour leur donner une signification.
- d) *Invention d'une méthode de mesure* : la démarche d'invention suppose, au-delà de l'observation et de l'interprétation des phénomènes, la sélection des variables pertinentes permettant la mise en place d'un modèle général.

Nous illustrerons cette étape maternelle, et pour la *descrip*

a) LANGUE MATERNELLE : *pertin*

Deux questions permettent d

- L'élève n'a-t-il pas introduit dans les textes mis à sa disposition toutes les informations qu'il a utilisées ?
- L'élève intègre-t-il les informations ?

Ces deux aspects à évaluer s

1° *L'introduction d'idées non pertinentes*

Les idées non pertinentes sont des informations apportées par les copies d'élèves a permis d'en évaluer l'aspect de la pertinence sémantique.

1. Le début du paragraphe 2 ne commence pas par « pousser la porte, entrer, ... » sont mentionnées dans le paragraphe 1).
2. Aucune allusion n'est faite au fait de pousser la porte, à une exclamation, à un seul mot.
3. Dans le paragraphe 2, le contenu du message est « l'élève devient une lettre, longue et ... ».
4. L'aspect et le contenu du message est « l'élève devient une lettre, longue et ... ».
5. L'enfant évoque des blessures.

2° *L'intégration des informations*

La seule estimation du non pertinent est de rendre compte de la valeur du message en appréciant la manière dont il est intégré dans le contexte plus large.

- a) On cherchera les indices de ce que l'élève a intégré :
 - l'évocation de la personne
 - les allusions à des personnes
 - l'expression des sentiments
 (Cette liste n'est pas exhaustive)
- b) On considérera que l'élève a intégré les informations si, dans son récit, il relate des faits qui se situent dans le 1^{er} paragraphe) et la réponse

DEUXIÈME ÉTAPE

Sélection et définition des aspects de chaque critère

Nous illustrerons cette étape pour la *pertinence sémantique*, en langue maternelle, et pour la *description de l'expérience*, en sciences.

a) LANGUE MATERNELLE : *pertinence sémantique*.

Deux questions permettent d'évaluer la pertinence sémantique :

- L'élève n'a-t-il pas introduit d'idées contredisant les informations contenues dans les textes mis à sa disposition ? Le travail ne doit pas nécessairement utiliser toutes les informations fournies. L'important est qu'il ne les contredise pas.
- L'élève intègre-t-il les informations retenues dans un contexte plus large ?

Ces deux aspects à évaluer sont décrits ci-après.

1° *L'introduction d'idées non pertinentes*.

Les idées non pertinentes sont celles qui entrent en contradiction avec les informations apportées par le texte. La lecture d'un nombre important de copies d'élèves a permis d'en établir la liste. Elle servira à évaluer ce premier aspect de la pertinence sémantique.

1. Le début du paragraphe 2 ne situe pas les faits dans le laboratoire (Exemple : pousser la porte, entrer,... sont des actions non pertinentes; elles sont déjà exprimées dans le paragraphe 1).
2. Aucune allusion n'est faite au « spectacle insolite ». Cette allusion peut se limiter à une exclamation, à un seul mot.
3. Dans le paragraphe 2, le commissaire est présent ou attendu.
4. L'aspect et le contenu du message sont modifiés (par exemple, le bout de papier devient une lettre, longue et soignée).
5. L'enfant évoque des blessures graves, des sinistres, des catastrophes,...

2° *L'intégration des informations*.

La seule estimation du nombre d'éléments non pertinents ne suffit pas à rendre compte de la valeur du travail de l'élève. On nuance cette estimation en appréciant la manière dont l'élève relate les faits en les intégrant dans un contexte plus large.

a) On cherchera les indices de cette intégration dans :

- l'évocation de la personnalité de l'oncle;
 - les allusions à des personnages secondaires;
 - l'expression des sentiments de Luc...
- (Cette liste n'est pas exhaustive).

b) On considérera que l'élève n'a pas intégré les informations s'il se contente de relater des faits qui se situent après l'entrée de Luc dans le laboratoire (fin du 1^{er} paragraphe) et la réponse de Luc au commissaire (début du 3^e paragraphe).

c) Dans certains travaux, on ne trouve que des idées qui, sans être nécessairement en contradiction avec les paragraphes 1 et 3, n'ont cependant pas de rapport avec la situation.

b) SCIENCES : description de l'expérience.

Trois aspects permettent d'apprécier la description écrite de l'expérience :

- Le caractère complet du travail.
- La précision des informations.
- Le respect de la chronologie.

A. Pour être complète, la description doit comporter les éléments suivants :

- A1. Égalité des volumes d'eau distillée de départ.
- A2. Inégalité des quantités de sel de calcium ajouté aux deux volumes d'eau.
- A3. Différence de limpidité entre les contenus des trois éprouvettes.
- A4. Inégalité des doses de savon conduisant à l'apparition d'une mousse persistante.

B. Pour être précise, la description doit contenir les informations suivantes :

- B1. Valeurs relatives des quantités de sel de Ca
(Exemple : simple - double, 1 dose - 2 doses, ou 1 g - 2 g).
- B2. Valeurs relatives des doses de savon
(Exemple : 6 mesures - 13 mesures).

C. La suite chronologique des phénomènes est respectée quand l'élève rend compte des faits dans l'ordre où ils se sont déroulés.

D'une manière générale, la sélection définitive des aspects de chaque critère est le fruit de mises au point successives, associées à la correction d'un grand nombre de travaux. Ainsi, pour l'évaluation de la pertinence sémantique, les professeurs souhaitaient d'abord valoriser l'introduction du plus grand nombre possible de références aux paragraphes d'encadrement. Mais la correction de travaux d'élèves a montré que leur qualité n'est pas nécessairement proportionnelle au nombre de références introduites. Pénaliser l'apparition des idées non pertinentes est plus significatif.

Alors que les critères doivent être indépendants, les aspects d'un critère sont, eux, nécessairement en relation.

Dans les échelles présentées, l'évaluation selon les différents critères résulte, soit de l'intervention de différents critères agissant comme des filtres successifs, soit de l'addition d'éléments juxtaposés. Par exemple, pour la pertinence sémantique, on évalue l'intégration des seuls éléments pertinents du travail. On peut ainsi nuancer l'évaluation portée sur des travaux équivalents pour le premier aspect. En sciences, la même relation unit les aspects « complet » et « précis » : on repère d'abord

des éléments significatifs de la situation avec laquelle certains d'entre eux ont précédé de la chronologie des phénomènes précédents, quelle que soit l'expérience.

Pour permettre un degré d'objectivité aux évaluateurs, la définition des aspects à évaluer est la plus précise possible.

Jusqu'ici, on s'est limité à une définition de la pertinence. Cette définition ne suffit cependant pas à évaluer encore spécifiquement les comportements des élèves par rapport à ces critères.

COMMENT IDENTIFIER LES COMPOSANTS DE LA PRODUCTION

Selon la nature de l'aspect à évaluer, les critères qui relèvent d'un univers de référence sont différents.

Dans le premier cas (univers de référence scientifique), les objectifs de l'exercice et les critères de correction permettent de dresser la liste des aspects à évaluer. Exemples : Éléments non pertinents, manque de précision, manque de complet et précis en sciences.

Dans le deuxième cas (univers de référence grammaticale), la liste complète des composants de la production doit prévoir tout ce que l'élève produit. Exemples : correction, de la précision, de la pertinence, de la structure grammaticales, par exemple.

Ici, il appartiendra donc à l'évaluateur de sélectionner la production à évaluer, les aspects à évaluer et les critères de correction. Cette situation est contraire à une objectivation de la production et à une unanimité des juges.

Afin de donner des points de référence, d'accord, on peut proposer des exemples de comportements à évaluer, ce que l'on fait dans les documents.

ni, sans être nécessairement en rapport avec la

tion écrite de l'expérience :

éléments suivants :

aux deux volumes d'eau.
deux éprouvettes.
l'apparition d'une mousse

Informations suivantes :

1 g - 2 g).

é quand l'élève rend compte

de des aspects de chaque
associées à la correction
évaluation de la pertinence
d valoriser l'introduction
aux paragraphes d'enca-
ves a montré que leur
au nombre de références
on pertinentes est plus

nts, les aspects d'un cri-

on les différents critères
res agissant comme des
 juxtaposés. Par exemple,
tégration des seuls élé-
ncer l'évaluation portée
ct. En sciences, la même
s » : on repère d'abord

des éléments significatifs de la description avant d'apprécier la précision avec laquelle certains d'entre eux sont relatés. Par contre, le respect de la chronologie des phénomènes est un critère qui s'ajoute aux deux précédents, quelle que soit l'évaluation portée en fonction de ceux-ci.

Pour permettre un degré d'accord élevé, voire l'unanimité des évaluateurs, la définition des aspects doit être aussi opérationnelle que possible.

Jusqu'ici, on s'est limité à présenter les critères retenus. Ce seul effort de définition ne suffit cependant pas à optimiser l'évaluation. Il faut encore spécifier les comportements qui traduisent le plus clairement ces critères.

COMMENT IDENTIFIER LES COMPORTEMENTS-CRITÈRES ?

Selon la nature de l'aspect à évaluer, les comportements qui l'actualisent relèvent d'un *univers fini* ou *non fini*.

Dans le premier cas (univers fini), les contraintes imposées par les objectifs de l'exercice et précisées par la formulation de la question permettent de dresser la liste exhaustive des comportements possibles. Exemples : Éléments non pertinents en langue maternelle — Caractère complet et précis en sciences.

Dans le deuxième cas (univers non fini), il est impossible de dresser la liste complète des comportements attendus parce qu'on ne peut prévoir tout ce que l'élève produira. Il en va ainsi pour l'appréciation de la correction, de la précision, de l'originalité du vocabulaire et des structures grammaticales, par exemple.

Ici, il appartiendra donc au professeur de recenser, pour chaque production à évaluer, les manifestations caractéristiques de l'aspect considéré. Cette situation est évidemment moins favorable que la première à une objectivation complète des critères, et il est vain d'espérer une unanimité des juges.

Afin de donner des points de repère et d'accroître ainsi les chances d'accord, on peut proposer, en pareil cas, une liste (non exhaustive) d'exemples de comportements, considérés comme caractéristiques. C'est ce que l'on fait dans les dossiers contenus dans la présente publication.

III. — TROISIÈME ÉTAPE

Détermination des niveaux d'exigence pour chaque aspect

Cette démarche comporte deux phases :

- Classer les comportements possibles en catégories hiérarchisées;
- Cerner les niveaux d'exigences convenant à un âge donné.

COMMENT CLASSER LES COMPORTEMENTS POSSIBLES EN CATÉGORIES HIÉRARCHISÉES ?

Dans le cas d'univers comportementaux finis, les caractéristiques de la matière conduisent à un choix. Ou bien on décide de considérer tous les comportements comme des unités équivalentes, ou bien on leur accorde une importance différente.

Si l'on décide de travailler avec des unités équivalentes, les distinctions entre échelons sont alors quantitatives et leur hiérarchie sera fondée sur une fréquence d'apparition. Par exemple, l'aspect « Introduction d'éléments non pertinents » relève de cette catégorie dans l'échelle « Pertinence sémantique ». Les contraintes imposées par les paragraphes d'encadrement ont, en effet, permis d'établir une liste exhaustive des éléments non pertinents possibles. Ils ont été considérés comme d'égale importance.

Si l'on pondère différemment les comportements retenus, les distinctions entre échelons revêtent alors un caractère qualitatif et la hiérarchie sera fondée sur l'importance relative des comportements en fonction des objectifs du travail. La définition du caractère complet dans l'échelle « Relation de l'expérience par une description écrite » (Sciences) relève de cette catégorie. Les caractéristiques de l'expérience proposée ont permis d'établir une liste de quatre éléments à relever dans la relation. On accorde le plus d'importance à la restitution des éléments codés A2 et A4. La description des échelons fait clairement apparaître cette pondération.

Dans le cas des univers comportementaux non finis, les unités répertoriées sont le plus souvent considérées comme équivalentes. Les distinctions entre échelons sont donc quantitatives et leur hiérarchie sera

fondée sur une fréquence d'apparition, la correction et de la précision du nombre d'incorrections et

REMARQUE.

Pour faire l'économie d'un échelon, on a parfois choisi de réduire le nombre de dichotomies ou trichotomies proches, mais cela n'est pas très satisfaisant, surtout si l'on a consacré à l'analyse. Ces cas

EXEMPLES :

1. *Univers fini* : Même poids à tous les échelons - 2^e cycle - Relation de l'expérience
— Pour le premier aspect « Caractère de la relation » retiendra cinq niveaux :
 - . Aucune erreur ou omission
 - . Une erreur ou une omission
 - . Deux erreurs ou deux omissions
 - . Plus de deux erreurs ou omissions
2. *Univers fini* : Pondération des échelons - 1^{er} cycle - Relation de l'expérience
— Pour l'aspect « Caractère de la relation »
 1. Les informations A1, A2 et A3
 2. Les informations A2 et A3
 3. L'information A2 ou A3
3. *Univers non fini* : Même poids à tous les échelons - 1^{er} cycle - Vocabulaire
— Pour l'aspect « Correction »
 1. Pas d'incorrection.
 2. Une ou deux incorrections.
 3. Plus de deux incorrections.
4. *Univers non fini* : Dichotomie des échelons - 1^{er} cycle - Pertinence sémantique
— Pour l'aspect « Intégration »
 1. Intégration.
 2. Pas d'intégration.
 3. Rien que des idées sa

pour chaque aspect

catégories hiérarchisées;

un âge donné.

CLASSIFIQUES EN CATÉGORIES

les caractéristiques de
décide de considérer tous
lentes, ou bien on leur

équivalentes, les distinc-
et leur hiérarchie sera
exemple, l'aspect « Intro-
te catégorie dans l'échelle
posées par les paragraphes
une liste exhaustive des
considérés comme d'égalé

ements retenus, les dis-
ctère qualitatif et la hié-
des comportements en
n du caractère complet
une description écrite »
ristiques de l'expérience
 quatre éléments à relever
ance à la restitution des
échelons fait clairement

on finis, les unités réper-
e équivalentes. Les dis-
s et leur hiérarchie sera

fondée sur une fréquence d'apparition. Par exemple, l'appréciation de la correction et de la précision du vocabulaire passe par l'identification du nombre d'incorrections et d'imprécisions observées dans le travail.

REMARQUE.

Pour faire l'économie d'un dénombrement fastidieux, nous avons parfois choisi de réduire le nombre de catégories à de simples dichotomies ou trichotomies proches de la paraphrase. Cette solution n'est cependant pas très satisfaisante. On en verra les lacunes dans le chapitre consacré à l'analyse. Ces cas sont repris dans la 4^e colonne du tableau.

EXEMPLES :

1. *Univers fini* : Même poids accordé à tous les comportements (Sciences - 2^e cycle - Relation de l'expérience).
— Pour le premier aspect « Relation des phénomènes essentiels », on retiendra cinq niveaux :
 - . Aucune erreur ou omission.
 - . Une erreur ou une omission.
 - . Deux erreurs ou deux omissions.
 - . Plus de deux erreurs ou omissions, ou aucune référence à l'une des deux phases.
2. *Univers fini* : Pondération variable selon les comportements (Sciences - 1^{er} cycle - Relation de l'expérience par une description écrite).
— Pour l'aspect « Caractère complet », on retiendra trois niveaux :
 1. Les informations A1, A2, A3 et A4 sont présentes.
 2. Les informations A2 et A4 au moins sont présentes.
 3. L'information A2 ou l'information A4 manque.
3. *Univers non fini* : Même poids accordé à tous les comportements (Langue maternelle - 1^{er} cycle - Vocabulaire).
— Pour l'aspect « Correction », on retiendra trois niveaux :
 1. Pas d'incorrection.
 2. Une ou deux incorrections.
 3. Plus de deux incorrections.
4. *Univers non fini* : Dichotomies ou trichotomies (Langue maternelle - 1^{er} cycle - Pertinence sémantique).
— Pour l'aspect « intégration », on retiendra trois niveaux :
 1. Intégration.
 2. Pas d'intégration.
 3. Rien que des idées sans rapport avec le contexte.

Langue maternelle	Univers fini		Univers non finis	
	Même poids accordé à tous les comportements; les comportements;	Même poids accordé à tous les comportements; les comportements;	Même poids accordé à tous les comportements;	Voir remarque p. 95
	Différences quantitatives entre échelons	Différences qualitatives entre échelons	Différences quantitatives entre échelons	
1 ^{er} cycle	1	2	3	4
Pertinence sémantique	Introduction d'éléments non pertinents			Intégration des informations
Pertinence syntaxique			Respect des temps; Respect des personnes grammaticales	
Vocabulaire			Correction; Précision	Originalité
Structures grammaticales			Correction; Variété	
Organisation des idées dans le paragraphe 2				- Cohérence sémantique - Liaison des idées - Progression du récit
2 ^e cycle				
Résumé	Redondances; idées secondaires	Compréhension des idées principales		
Critique		Nombre et correction des idées reprises; Validité et type d'arguments		

Sciences	Univers fini		Univers non finis	
	Même poids accordé à tous les comportements; les comportements;	Même poids accordé à tous les comportements; les comportements;	Même poids accordé à tous les comportements;	voir remarque p. 95
	Différences quantitatives entre échelons	Différences qualitatives entre échelons	Différences quantitatives entre échelons	
1 ^{er} cycle	1	2	3	4

2 ^e cycle				
Résumé	Redondances: idées secondaires	Compréhension des idées principales		
Critique		Nombre et correction des idées reprises; Validité et type d'arguments		

Sciences	Univers fini / Univers non finis			
	1	2	3	4
1 ^{er} cycle				
Relation de l'expérience par une description écrite	Précision	Caractère complet		Respect de la chronologie
Relation de l'expérience par un schéma	Précision	Caractère complet		Caractère dynamique
Interprétation	Caractère complet	Précision		
Invention d'une méthode de mesure		Invention		
2 ^e cycle				
Relation de l'expérience	- Relation des phénomènes essentiels; - Distinction-localisation-répétition			
Interprétation des phénomènes observés	- Classement des halogènes - Justification du classement	Identification du corps. Distinction entre phénomènes physiques et réactions chimiques		
Invention d'une méthode d'analyse		Invention		

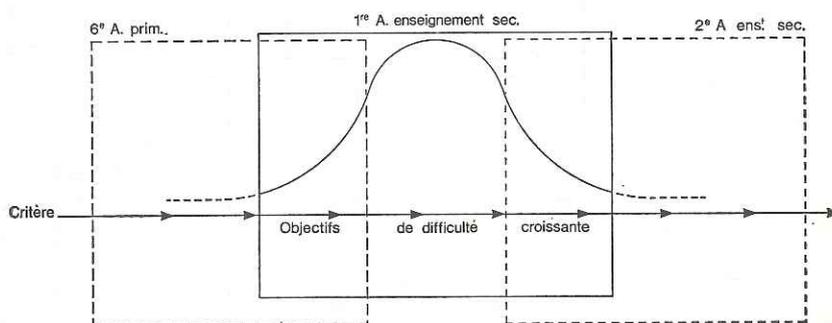
COMMENT CERNER LES NIVEAUX D'EXIGENCE CONVENANT À UN ÂGE DONNÉ ?

Pour cerner les exigences qui correspondent à un niveau pédagogique donné, on se fonde sur l'analyse d'un grand nombre de travaux d'élèves appartenant, non seulement au niveau pédagogique choisi, mais aussi aux niveaux immédiatement inférieur et supérieur ⁽¹⁾. Ainsi, pour l'expérience menée en première année de l'enseignement secondaire rénové, on a rassemblé des travaux de sixième année primaire et de deuxième année de l'enseignement secondaire rénové.

Les informations recueillies par l'analyse du corpus permettent :

- de déterminer, dans la gamme des comportements caractéristiques d'un aspect, les exigences minimales et maximales pour un niveau pédagogique donné;
- de regrouper éventuellement les comportements compris dans ces limites.

La combinaison de ces deux démarches doit assurer une allure gaussienne à la répartition des évaluations, comme l'illustre le schéma suivant :



Un critère d'évaluation est choisi parce qu'il correspond à un objectif éducatif jugé important. Lorsqu'il s'agit d'apprentissages complexes tels que la capacité de s'exprimer clairement par écrit, l'objectif devient

⁽¹⁾ Pour garantir la validité de l'analyse du corpus, on s'efforce de constituer un échantillon aussi représentatif que possible de la population scolaire.

lui aussi de plus en plus co
et le niveau d'exigence de l'

La figure ci-dessus schém

Le cadre en trait plein co
dération en première année

Mal adapter les exigences
trop peu discriminative; e
serait concentrée dans les é
rité des comportements ser

IV.

De

Afin de préciser au mi
jusqu'à présent, adopté un
des *critères*, on s'est efforc
cun d'eux et de préciser, c
à retenir. On a vu aussi qu
complément; pour être sig

Selon le nombre d'aspe
marche conduit à un nomb
décider de considérer certai
et ainsi aboutir à la même
l'on a déterminé l'import

Si l'on décide d'accorde
niveaux de chacun d'eux in
parfaitement. C'est le cas,
de l'échelle « Organisation

Cohérence sémantique Liaison

- | | | |
|-----------------------------|---|-----------|
| — Récit cohérent | } | — Liaison |
| | | — Liaison |
| — Contradictions passagères | } | — Liaison |
| | | — Liaison |

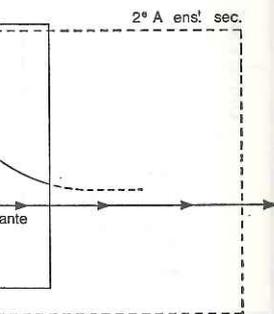
CONVENANT À UN ÂGE

un niveau pédagogique
nombre de travaux d'élèves
technique choisi, mais aussi
pour (1). Ainsi, pour l'ex-
piment secondaire rénové,
primaire et de deuxième

le corps permettent :
caractéristiques
maximales pour un niveau

contenus compris dans ces

assurant une allure gaus-
sienne l'illustre le schéma



correspond à un objectif
d'apprentissages complexes
écrit, l'objectif devient

on s'efforce de constituer un
relation scolaire.

lui aussi de plus en plus complexe à mesure que les études progressent et le niveau d'exigence de l'évaluateur s'élève parallèlement.

La figure ci-dessus schématise l'évolution de l'évaluation.

Le cadre en trait plein correspond aux comportements pris en considération en première année de l'enseignement rénové.

Mal adapter les exigences au niveau des élèves conduirait à une échelle trop peu discriminative; elle serait soit trop indulgente (la majorité serait concentrée dans les échelons supérieurs), soit trop sévère (la majorité des comportements serait concentrée dans les échelons inférieurs).

IV. — QUATRIÈME ÉTAPE

Description des échelons

Afin de préciser au mieux les composantes de l'évaluation, on a, jusqu'à présent, adopté une démarche analytique : après la sélection des *critères*, on s'est efforcé de déterminer les *aspects* essentiels de chacun d'eux et de préciser, dans chaque cas, les *types de comportements* à retenir. On a vu aussi que les différents aspects d'un même critère se complètent; pour être significative, la note devra donc les combiner.

Selon le nombre d'aspects retenus et les niveaux d'exigence, la démarche conduit à un nombre élevé de combinaisons. On peut cependant décider de considérer certaines de ces combinaisons comme équivalentes et ainsi aboutir à la même note. Cette décision suppose cependant que l'on a déterminé l'importance relative des différents aspects.

Si l'on décide d'accorder le même poids à tous les aspects, tous les niveaux de chacun d'eux interviendront dans l'échelle en se compensant parfaitement. C'est le cas, par exemple, pour les trois premiers niveaux de l'échelle « Organisation des idées » (1^{er} cycle).

<i>Cohérence sémantique</i>	<i>Liaison des idées</i>	<i>Progression du récit</i>	
— Récit cohérent	{ — Liaison adroite — Liaison maladroite	{ — Récit non redondant	TB
		{ — Progression ralentie	B
— Contradictions passagères	{ — Liaison adroite	{ — Récit non redondant	B
		{ — Progression ralentie	S
	{ — Liaison maladroite	{ — Récit non redondant	S
		{ — Progression ralentie	F

<i>Cohérence sémantique</i>	<i>Liaison des idées</i>	<i>Progression du récit</i>	
— Récit incohérent	— Liaison adroite	— Récit non redondant	F
		— Progression ralentie	I
	— Liaison maladroite	— Récit non redondant	I
		— Progression ralentie	I

On remarquera que la note B est attribuée dans trois cas :

- Récit cohérent — liaison adroite — progression ralentie
- Récit cohérent — liaison maladroite — récit non redondant
- Contradictions — liaison adroite — récit non redondant passagères

De cette façon, une lacune est pénalisée de la même manière dans chacun des trois aspects.

Par contre, si l'on veut attribuer moins d'importance à un aspect, on construira l'échelle de façon telle qu'il n'interviendra que pour nuancer les premiers niveaux d'un autre aspect.

EXEMPLE : échelle « Description écrite de l'expérience » (1^{er} cycle).

		<i>Respect de la chronologie</i>	
<i>Caractère complet</i>	<i>Précision</i>		
— A1 A2 A3 et A4	— B1 et B2	— Oui	TB
		— Non	B
	— B1 ou B2	— Oui	B
		— Non	S
	— Ni B1 ni B2	— Oui	S
		— Non	F
— A2 et A4	— B1 et B2	— Oui	B
		— Non	S
	— B1 ou B2	— Oui	S
		— Non	F
	— Ni B1 ni B2	— Oui	F
		— Non	I
— A2 ou A4	— Quelles que soient les caractéristiques de précision et de chronologie	I	

De même, on peut souhaiter pénaliser davantage l'apparition d'un type de réponse pour un aspect; c'est le cas, par exemple, pour le troisième niveau de l'aspect « Intégration » (échelle « Pertinence sémantique »).

Introduction d'éléments non pertinents

- Pas d'éléments non pertinents
- 1 ou 2 éléments non pertinents
- 3 ou 4 éléments non pertinents
- Plus de 4 éléments non pertinents

La pondération appliquée (en fonction de la situation), dans chaque niveau d'exigence, pénalise toute lacune dans...

Présentation de l'étape précédente

L'étape précédente combinée avec la situation précédente. Il reste à présenter l'ensemble des deux types de démarcation.

D'abord, on suit la première partie de l'étape précédente décrites jusqu'ici sont retenus sont justifiés. Ensuite, on présente les informations retenues pour analyser la situation. On examine, en premier lieu, les éléments les plus bas; on passe ensuite aux éléments les plus élevés. On trouvera des exemples de réponses qui figurent en annexe.

La seconde présentation des critères sont organisés en fonction d'une hiérarchie des compétences.

Expression du récit
 non redondant F
 Expression ralentie I
 non redondant I
 Expression ralentie I
 Dans trois cas :
 Expression ralentie
 Récit non redondant
 Récit non redondant

de la même manière dans
 l'importance à un aspect,
 l'interviendra que pour
 cet.
 « (1^{er} cycle).

..... TB
 B
 B
 S
 S
 F
 B
 S
 S
 F
 F
 I
 Caractéristiques de précision
 I

l'apparition d'un
 exemple, pour le troi-
 « Pertinence sémantique

Introduction d'éléments non pertinents

Intégration

— Pas d'éléments non pertinents	}	— Intégration	TB
		— Pas d'intégration	B
		— Uniquement des idées sans rapport direct avec la situation	I
— 1 ou 2 éléments non pertinents	}	— Intégration	B
		— Pas d'intégration	S
		— Uniquement des idées sans rapport direct avec la situation	I
— 3 ou 4 éléments non pertinents	}	— Uniquement des idées sans rapport direct avec la situation	I
— Plus de 4 éléments non pertinents		I

La pondération appliquée au seul niveau inférieur (idées sans rapport avec la situation), dans cet exemple, pourrait être généralisée pour chaque niveau d'exigence. Autrement dit, on pénaliserait alors davantage toute lacune dans le deuxième aspect.

CINQUIÈME ÉTAPE

Présentation de l'échelle et démarche pour son utilisation

L'étape précédente a permis d'énoncer les règles qui président à la combinaison des critères établis pour chacun des aspects de l'évaluation. Il reste à présenter l'ensemble des critères sous la forme la plus adéquate. Deux types de démarches ont été retenus.

D'abord, on suit la démarche chronologique de l'évaluateur. Dans une première partie du dossier, les informations relatives aux étapes décrites jusqu'ici sont abordées d'un point de vue théorique et les critères retenus sont justifiés. Pour faciliter l'évaluation, on reprend systématiquement les informations essentielles dont le professeur devra tenir compte pour analyser la performance de l'élève : pour chaque aspect, on examine, en premier lieu, les niveaux d'exigence, du plus élevé au plus bas; on passe ensuite à la synthèse des caractéristiques repérées. On trouvera des exemples de cette démarche avant chacune des échelles qui figurent en annexe.

La seconde présentation n'apporte pas d'informations nouvelles. Les critères sont organisés de façon différente, afin de faire apparaître la hiérarchie des comportements relatifs à chaque dimension. A cet effet,

on réunit en un tableau les caractéristiques qui conduisent à une même note. Cette présentation permet, en outre, de rendre plus complète l'information donnée à l'élève sur la valeur de la performance : elle indique les étapes qui restent à parcourir pour maîtriser les comportements visés.

Pour être opérationnelle, l'échelle d'évaluation doit éviter deux écueils :

- La paraphrase pure et simple des catégories ou la définition de critères en termes mentalistes.
- La recherche d'un instrument totalement objectif, mais impraticable dans la vie quotidienne d'une classe.

Pour illustrer le premier écueil, nous reviendrons sur un exemple déjà cité, l'échelle d'évaluation des compositions construites par l'*Educational Testing Service* de Princeton (1). Pour évaluer les idées contenues dans des compositions. Voici les échelons retenus :

IDÉES (contenu).

— Niveau élevé.

L'étudiant traite de tous les points appelés par le sujet de la composition ou par le plan du 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.

— 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.

L'effort de définition des critères présentés ci-dessus ne manque pas d'intérêt. On peut cependant craindre que leur utilisation conduise à bien des discordances encore. Le lecteur manque d'informations : quels

(1) E.T.S., *Definition of Ratings on the ETS Composition Scale, o.c.*, p. 3.

sont les points appelés un esprit richement meublé. Le travail se fasse en fonction de ce qu'il faut-il pas, au contraire, il faut juger de la valeur d'une

Le souci d'une plus grande précision dans l'élaboration d'instruments qui nécessitent de longues études, par exemple, qui ont été analysés par référence à des compositions peuvent être menés par étapes, commencé par mettre à jour certains aspects. Pour la sémantique, nous avons utilisé les informations, en faisant appel à deux listes de contraintes (x2), les premières de 10 secondes. Ensuite, on a demandé à l'élève avait tenu compte

Toutes ces informations ont été prises en compte dans la sémantique :

Chaque degré de l'échelle

Un tel instrument qui est d'une utilisation plus précise dans le dossier « économique pour lequel seule la liste des éléments pour évaluer l'intégration des éléments frustes, mais précisées

sont les points appelés par le sujet, qu'est-ce qu'une définition claire, un esprit richement meublé?... De plus, faut-il que l'évaluation d'un travail se fasse en fonction de l'ensemble des travaux de la classe? Ne faut-il pas, au contraire, que les critères d'évaluation permettent de juger de la valeur d'une performance en elle-même?

Le souci d'une plus grande objectivité peut conduire le chercheur à l'élaboration d'instruments sophistiqués, nés d'études fondamentales et qui nécessitent de longs et minutieux dépouillements. On pourrait imaginer, par exemple, que le vocabulaire de chaque composition soit analysé par référence à des vocabulaires de base. De tels dépouillements peuvent être menés par ordinateur. Dans cet ordre d'idées, nous avons commencé par mettre au point des indices numériques pour l'évaluation de certains aspects. Par exemple, pour l'évaluation de la pertinence sémantique, nous avons tenté d'apprécier le niveau d'intégration des informations, en faisant d'abord le recensement de toutes les informations fournies par les paragraphes d'encadrement, ce qui permettait de dresser deux listes de contraintes, les unes obligatoires (x_1) et les autres facultatives (x_2), les premières se voyant réserver plus d'importance que les secondes. Ensuite, on comptait le nombre d'éléments (x_1 et x_2) dont l'élève avait tenu compte, puis le nombre d'éléments non pertinents (y).

Toutes ces informations étaient synthétisées en une formule « Pertinence sémantique » :

$$PS = \frac{2x_1 + x_2}{y}$$

Chaque degré de l'échelle était numériquement délimité.

Un tel instrument garantit une objectivité quasi totale. Toutefois, il est d'une utilisation peu commode dans la pratique courante. On trouvera dans le dossier « Langue maternelle, 1^{er} cycle », la solution plus économique pour laquelle nous avons finalement opté. On verra que seule la liste des éléments non pertinents a été maintenue, tandis que, pour évaluer l'intégration, on se contente de quelques catégories assez frustes, mais précisées par des exemples.

B. — REMANIEMENT DE L'INSTRUMENT

I. — RECHERCHE DES VICES DE CONSTRUCTION ET D'UTILISATION

L'application de la méthode décrite ci-dessus a conduit à une première rédaction des instruments. Un contrôle de validité a alors été entrepris, la question étant : « Les échelles descriptives permettent-elles une évaluation plus objective et plus nuancée que les échelles habituellement utilisées ? » Seule une expérience menée dans des conditions strictement contrôlées permet de répondre à pareille question.

Avant de prendre contact avec un grand nombre d'enseignants, nous avons soumis le dossier à quelques professeurs afin de connaître leurs premières réactions. Cette démarche préliminaire permet généralement d'éliminer, avant l'expérimentation massive, les ambiguïtés les plus graves et les erreurs les plus grossières.

On trouvera au chapitre suivant les résultats de l'expérimentation. Sans entrer maintenant dans les détails de ces résultats, on constatera qu'ils sont loin de refléter une situation parfaite. Une double hypothèse s'offrirait pour expliquer la concordance, encore relativement peu élevée entre les juges : des erreurs de lecture et de manipulation de la part des utilisateurs, et des erreurs de construction pour certaines échelles.

a) MÉTHODE.

Dans cette phase de la recherche, on a recouru à la méthode clinique, afin de découvrir la démarche du professeur qui évalue en fonction des critères proposés. La technique de la réflexion parlée s'avère fort utile en pareil cas ; cette technique est définie comme suit par G. De Landsheere : « Le sujet est invité à formuler à haute voix les démarches de sa pensée, pendant qu'il résout un problème, ce qui permet d'étudier les processus mentaux dans leur déroulement et donc d'identifier les causes des succès et des erreurs » (1).

Cette démarche analytique est lourde, mais la richesse des informations de caractère diagnostique qu'elle apporte permet de limiter le nombre de

(1) G. DE LANDSHEERE, *Introduction à la recherche en éducation*, 4^e éd., o.c., p. 82.

professeurs interviewés. s'est prêté à l'expérience mant, à voix haute, ses

Pour surmonter les difficultés posés dans le dossier, ont été sollicités.

Dans un second temps, afin d'obtenir une réponse résultats. Voici quelques (technique) :

- Indiquez, dans le travail, quel est le processus généralement penser qu'un inventées.
- Certains professeurs (44) Quelle différence faites-vous ?
- Qu'est-ce qui vous permet de ne pas confondre ?

Pour l'attribution de la note globale, on pose ce questionnaire au professeur est simplement inviter aux appréciations analytiques. Pour la note globale, on pose ce questionnaire méthode utilisée :

1. Comment procédez-vous pour la note globale ?
2. Avez-vous attribué une note globale ? Oui - Non.
3. Si vous n'avez pas accordé la note globale, quel est le critère qui a déterminé votre décision ? (en soi, ou le degré de difficulté) ?

b) DÉPOUILLEMENT.

Les informations recueillies sont classées en deux types d'observations :

- Les comportements des élèves
- Les vices de construction

INSTRUMENT

CONSTRUCTION ET D'UTILISATION

... a conduit à une pre-
de validité a alors été
riptives permettent-elles
e les échelles habituelle-
ans des conditions stric-
e question.

nombre d'enseignants,
seurs afin de connaître
minaire permet généra-
massive, les ambiguïtés

s de l'expérimentation.
résultats, on constatera
Une double hypothèse
relativement peu élevée
ipulation de la part des
ur certaines échelles.

u à la méthode clinique,
évalue en fonction des
parlée s'avère fort utile
t par G. De Landsheere :
démarches de sa pensée,
d'étudier les processus
ier les causes des succès

chesse des informations
de limiter le nombre de

ducation, 4^e éd., o.c., p. 82.

professeurs interviewés. En un premier temps, chaque professeur qui s'est prêté à l'expérience a été invité à noter le travail présenté en exprimant, à voix haute, ses hésitations, ses choix et les raisons de ceux-ci.

Pour surmonter les difficultés d'intégration de tous les critères proposés dans le dossier, seuls des professeurs du groupe expérimental ont été sollicités.

Dans un second temps, on a procédé à des interviews plus structurées afin d'obtenir une réponse à des questions suggérées par l'analyse des résultats. Voici quelques exemples de questions posées (travail scientifique) :

- Indiquez, dans le travail de l'élève, la ou les phrases qui vous font éventuellement penser qu'une ou plusieurs méthodes de mesure ont bien été inventées.
- Certains professeurs (44 %) ont attribué la note B. A votre avis, pourquoi ? Quelle différence faites-vous entre les échelons TB et B ?
- Qu'est-ce qui vous permet de dire que le travail de l'élève établit ou n'établit pas la distinction entre les deux moments de l'expérience.

Pour l'attribution de la note globale, la question est ouverte. Le professeur est simplement invité à évaluer; on ne suggère pas qu'il se réfère aux appréciations analytiques précédentes. Après l'attribution de la note globale, on pose cependant des questions destinées à révéler la méthode utilisée :

1. Comment procédez-vous pour passer des appréciations analytiques à l'appréciation globale ?
2. Avez-vous attribué une importance égale à chacun des critères de l'évaluation ? Oui - Non.
3. Si vous n'avez pas accordé le même poids aux différents critères, qu'est-ce qui a déterminé votre décision : l'importance relative de chaque critère en soi, ou le degré de difficulté que les différents aspects considérés présentent pour des élèves du niveau scolaire auquel vous enseignez ?

b) DÉPOUILLEMENT.

Les informations recueillies ont été consignées sur des fiches relatives à deux types d'observation :

- Les comportements du professeur vis-à-vis de l'instrument.
- Les vices de construction affectant les échelles.

c) RÉSULTATS DES INTERVIEWS.

1° Erreurs d'utilisation.

a) Erreurs de manipulations.

Certains comportements témoignent du désir, chez certains professeurs, de se libérer plus ou moins des contraintes proposées, pour évaluer de façon plus personnelle :

- Accorder une note, pour chaque critère, sans utiliser l'échelle.
- Accorder une note, pour chaque critère, de manière empirique, et la justifier, en cherchant dans l'échelle une description qui puisse apporter une justification *a posteriori*.
- Utiliser la description des aspects pour se faire une opinion sur le travail, sans évaluer ces aspects selon les directives fournies.
- Accorder une note, pour un critère, en ne tenant compte que d'un aspect.
- Utiliser l'échelle sans avoir lu la description préalable des aspects.
- Introduire des critères qui ne figurent pas dans les dossiers.
- Introduire des critères qui visent à nuancer la note en fonction du moment de l'année auquel le travail a été proposé aux élèves, ou en fonction du niveau de la classe auquel le professeur s'adresse généralement.
- Introduire des critères relatifs à la matière : ajouter des critères nouveaux (orthographe, ponctuation); ajouter des aspects nouveaux dans des critères prévus (originalité, dans la pertinence sémantique).

b) Erreurs de lecture.

Ces erreurs se situent soit dans la compréhension des critères isolés, soit dans la combinaison des critères de l'échelle.

Dans l'échelle « Pertinence sémantique », par exemple, certains professeurs interprètent mal les consignes relatives à l'introduction d'éléments non pertinents. Ils pénalisent l'absence de certains faits présentés dans les paragraphes d'encadrement, au lieu de ne relever que les contradictions dont on a fourni la liste.

Pour l'évaluation de la dimension « Relation de l'expérience par une description écrite », en sciences, le professeur relève, par exemple, dans le travail de l'élève, les caractéristiques A2 et A4, B1 et B2, et considère que la chronologie est respectée. Il attribue ensuite la note B, alors que la combinaison de ces qualités devrait conduire à la note S.

Ce relevé des erreurs de manipulation et de lecture ne justifie pas une attitude pessimiste. Les conditions de la recherche les expliquent en partie. Les dossiers, du moins tels qu'ils ont été présentés dans leur première version, apparaissent d'une grande complexité pour qui n'a

pas participé à leur élaboration et les professeurs artificielle peut les incliner à une notation trop lourde pour la tâche. De plus, il aurait fallu préciser sur les objectifs poursuivis l'efficacité de critères établis, indispensable que les manuels soient strictement aux normes, elles ne correspondaient pas, il aurait fallu inviter les participants, ils auraient souhaité ret...

2° Vices de construction

Bon nombre de défauts ont dû être évités si la méthode est strictement établie dès le début de l'expérimentation et les règles essentielles à respecter e...

Parmi les défauts dont on note :

- Dans certains cas, les notes. La première version c...

	Correction
TB	Correct
B	Le plus souvent correct
S	Généralement correct ou : Généralement correct
F	Généralement correct
I	Souvent incorrect

pas participé à leur élaboration. En effet, les critères proposés sont nombreux et les professeurs ne les utilisent qu'une seule fois. Cette situation artificielle peut les incliner à considérer que l'instrument est d'application trop lourde pour la notation de l'ensemble des travaux d'une classe. De plus, il aurait fallu apporter quelques précisions supplémentaires sur les objectifs poursuivis dans cette expérience. Il s'agissait de tester l'efficacité de critères établis par un groupe d'enseignants. Aussi était-il indispensable que les maîtres dont on sollicitait la collaboration se conforment strictement aux consignes de correction proposées, même si elles ne correspondaient pas à leurs propres options. Il aurait d'ailleurs fallu inviter les participants à l'expérience à préciser quels autres critères ils auraient souhaité retenir pour la notation de la performance.

2^o Vices de construction des échelles.

Bon nombre de défauts de construction relevés ci-après auraient pu être évités si la méthode décrite dans les pages qui précèdent avait été strictement établie dès le départ. On l'a déjà indiqué, c'est des résultats de l'expérimentation et de leur analyse que se sont dégagées les règles essentielles à respecter et les précautions à prendre.

Parmi les défauts dont la constatation offre un intérêt méthodologique, on note :

— Dans certains cas, les catégories distinguées n'étaient pas *exhaustives*. La première version de l'échelle « Vocabulaire » se présentait comme suit :

	Correction	Précision	Variété
TB	Correct	précis	et varié
B	Le plus souvent correct	précis	et varié
S	Généralement correct	imprécis	et varié
	ou : Généralement correct	généralement précis	et varié
F	Généralement correct	imprécis	mais peu varié
I	Souvent incorrect	imprécis	et peu varié

On remarquera notamment qu'une possibilité n'est pas prévue : vocabulaire correct, imprécis et varié.

— D'autres formulations étaient *ambiguës*. Dans l'échelle « Invention d'une méthode de mesure », la distinction entre les échelons S et I n'était pas suffisamment claire :

S : L'élève suggère une méthode appropriée au contexte, mais la présente de manière imprécise ou incohérente.

I : Aucune suggestion, ou la méthode suggérée est inadéquate.

L'hésitation entre les deux échelons s'est clairement traduite dans les histogrammes de fréquence des notes. Près de 80 % des professeurs répartissent leurs évaluations de manière quasi égale entre ces deux échelons.

L'échelle « Pertinence sémantique » présentait un autre cas d'ambiguïté. La définition de l'aspect « Intégration » était précisée par deux exemples. Les professeurs les ont considérés comme les éléments d'une liste exhaustive et en ont relevé la présence ou l'absence dans le travail de l'élève. Autrement dit, l'exemple présenté comme illustration du trait à évaluer a pris valeur de critère.

— On observe enfin une échelle dont les degrés (sous-catégories) n'étaient pas *mutuellement* exclusifs. Dans l'échelle « Invention d'une méthode de mesure », deux niveaux ne sont pas totalement indépendants :

S : Une méthode appropriée au contexte est suggérée, mais elle est présentée de manière imprécise ou incohérente.

F : L'élève suggère la méthode de l'évaporation.

Certains travaux d'élèves ont ébauché une méthode appropriée au contexte de l'expérience, puis ont suggéré la méthode de l'évaporation. Telle qu'elle était construite, l'échelle ne permettait pas de préciser s'il fallait ranger les réponses de ce type en S ou en F.

II. — COMMENT RÉDUIRE LES DISCORDANCES ENTRE LES JUGES ?

L'application de la méthode proposée dans les pages qui précèdent permet de pallier les vices de construction repérés dans les échelles. On s'efforcera de définir chacun des critères en termes de comportements attendus chez l'élève (en particulier, il importe de fixer clairement les niveaux d'exigence) et de supprimer ainsi l'ambiguïté des termes trop généraux.

On s'efforcera aussi d'ajouter des règles systématiques relatives à la combinaison des critères ainsi précisés. De cette manière, on respecte

les exigences fondamentales et l'indépendance des critères.

Pour limiter les erreurs :

1. clarifier les consignes et les *niveaux d'exigence* ;
2. programmer la démarche à se prononcer successivement sur les choix multiples et les possibilités présentées par chacun des aspects.

Le professeur est tenu de préciser sa démarche avant de formuler ses jugements. Cette démarche est, elle-même, à évaluer.

Le caractère analytique des critères doit permettre de retrouver aisément les enseignements devant être évalués.

pas prévue : vocabulaire

échelle « Invention d'une
niveaux S et I n'était pas

contexte, mais la présente

est inadéquate.

nt traduite dans les histo-
s professeurs répartissent
es deux échelons.

n autre cas d'ambiguïté-
cisée par deux exemples-
ents d'une liste exhaustive
travail de l'élève. Autre-
du trait à évaluer a pris

s-catégories) n'étaient pas
n d'une méthode de me-
dépendants :

gérée, mais elle est pré-

de appropriée au contexte
évaporation. Telle qu'elle
ciser s'il fallait ranger les

ENTRE LES JUGES ?

es pages qui précèdent
opérés dans les échelles.
rmes de comportements
de fixer clairement les
biguïté des termes trop

ématiques relatives à la
te manière, on respecte

les exigences fondamentales de toute analyse de contenu : l'exhaustivité et l'indépendance des catégories.

Pour limiter les erreurs de lecture, on s'est efforcé de :

1. clarifier les consignes en définissant systématiquement les *critères* et les *niveaux d'exigence*;
2. programmer la démarche d'évaluation. Le professeur est invité à se prononcer successivement sur *chacun* des aspects. La présentation en choix multiples oblige le professeur à tenir compte de toutes les possibilités présentées. L'obligation de répéter cette démarche pour chacun des aspects supprime le risque d'en négliger certains.

Le professeur est invité à combiner ses appréciations successives avant de formuler une évaluation relative à chaque critère. Cette démarche est, elle aussi, simplifiée par une présentation en réseaux.

Le caractère analytique de cette démarche permet, en outre, de retrouver aisément les sources de discordance entre juges. De tels renseignements devraient permettre des améliorations ultérieures.

RÉSULTATS EXPÉRIMENTAUX

Construire un instrument de mesure appliquant une méthode éprouvée permet d'espérer une bonne validité. Mais cet espoir doit être vérifié par les faits. Autrement dit, les instruments construits doivent subir l'épreuve de l'expérience.

Dans le présent chapitre, on rapporte les principales indications fournies par deux expériences successives. Le détail de l'expérimentation a été reporté en annexe.

I. — PREMIÈRE EXPÉRIMENTATION

a) SCHÉMA EXPÉRIMENTAL.

Des professeurs ont été invités à évaluer *un seul travail d'élève*. Cette situation peut paraître artificielle, car, dans la pratique quotidienne de la classe, un enseignant évalue presque toujours un grand nombre de copies et émet un jugement en partie relatif. Ici, on a voulu concentrer au maximum l'attention sur la matière.

Tous les professeurs de langue maternelle ou de sciences, engagés dans la rénovation au moment de l'expérience, ont été contactés. Ils ont été répartis en trois groupes.

- . Un *premier groupe* a évalué globalement le travail de l'élève à l'aide d'une simple échelle verbale à cinq degrés.
- . Un *deuxième groupe* a évalué le même exercice :
 - *analytiquement* d'abord, à l'aide d'échelles verbales relatives à plusieurs critères préalablement définis;
 - *globalement* ensuite, à l'aide d'une échelle verbale à cinq degrés.
- . Un *troisième groupe* a évalué le même exercice :
 - *analytiquement* d'abord, à l'aide d'échelles descriptives relatives aux mêmes critères;

— globalement ens

Groupe I (Groupe de contrôle)
(Empty space for data or notes)
Appréciation globale TB B S F I [] [] [] [] []

L'expérience tente

1. Le passage par d'cordance des app complexe? La co groupes I et II ap
2. La définition préc des appréciations dans les groupes II
3. Les justifications c aux objectifs du t d'évaluations desc

Les professeurs de appréciations analyt de ces justifications d'évaluation utilisés formation fournie à l

L'utilisation des éc rend superflue la just

— globalement ensuite, à l'aide d'une échelle verbale à cinq degrés.

Groupe I (Groupe de contrôle)	Groupe II (Groupe de contrôle)	Groupe III (Groupe expérimental)
	<i>Appréciations analytiques</i> Critère 1 : T B B S F I _ _ _ _ _ _ _ Critère 2 : T B B S F I _ _ _ _ _ _ _ Critère 3 : T B B S F I _ _ _ _ _ _ _ 	<i>Appréciations analytiques</i> Critère 1 : T B B S F I _ _ _ _ _ _ _ Critère 2 : T B B S F I _ _ _ _ _ _ _ Critère 3 : T B B S F I _ _ _ _ _ _ _ Chaque échelon est défini.
<i>Appréciation globale</i> T B B S F I _ _ _ _ _ _ _	<i>Appréciation globale</i> T B B S F I _ _ _ _ _ _ _	<i>Appréciation globale</i> T B B S F I _ _ _ _ _ _ _

L'expérience tente d'élucider trois questions.

1. Le passage par des appréciations analytiques augmente-t-il la concordance des appréciations globales portées sur une performance complexe ? La comparaison des notes globales octroyées dans les groupes I et II apportera des éléments de réponse à cette question.
2. La définition précise de chaque critère augmente-t-elle la concordance des appréciations analytiques ? La comparaison des notes attribuées dans les groupes II et III devrait permettre de répondre à cette question.
3. Les justifications des notes sont-elles plus précises et mieux adaptées aux objectifs du travail lorsque les professeurs utilisent des échelles d'évaluations descriptives ?

Les professeurs des groupes I et II ont été invités à justifier leurs appréciations analytiques et/ou globales. Nous voulions, par l'analyse de ces justifications, tester l'hypothèse de la spécificité des critères d'évaluation utilisés par le maître, d'une part, et de la spécificité de l'information fournie à l'élève sur la valeur de sa performance, d'autre part.

L'utilisation des échelles descriptives par les professeurs du groupe III rend superflue la justification de leurs appréciations analytiques. En effet,

le choix d'un niveau de l'échelle rend compte avec précision des qualités du travail de l'élève. Ici, seule la note globale doit être justifiée.

b) ÉCHANTILLON.

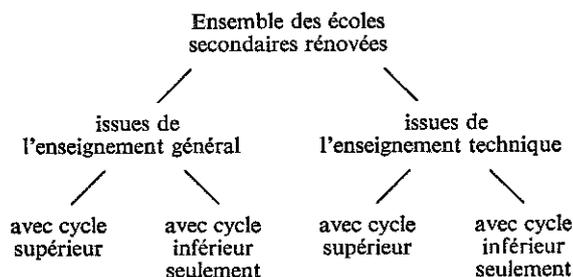
1° Critères d'échantillonnage.

Pour faciliter le déroulement de la recherche, les professeurs d'un même établissement ont été affectés à un même groupe. C'est pourquoi l'échantillonnage se fait par écoles.

On a choisi des critères d'échantillonnage dont on pouvait supposer qu'ils influencent la manière d'évaluer :

- Type d'enseignement auquel l'école appartenait avant la rénovation : général ou technique.
- Ecole de six années ou école ne comptant que le cycle inférieur (cette caractéristique est liée à la proportion de régents ou de licenciés enseignant au cycle inférieur dans l'établissement).
- Taille de l'établissement (nombre d'élèves dans l'école).

Ces trois facteurs ont été successivement pris en considération dans la constitution de l'échantillon. On peut représenter schématiquement les niveaux successifs de la stratification.



les écoles de 1 à n sont classées par ordre de grandeur	}	1	1	1	1
	
	
	
		n	n	n	n

Les trois groupes ainsi les trois critères d'échan

2° Grandeur de l'échanti

Pour la première expé seurs de langue matern dans les 132 écoles ayan

Pour ce genre d'entre été élevé : 36 % des professeurs de sciences.

II. —

Sur la base des résulta des remaniements ont é struction de l'instrumen

Les instruments rema tation. Celle-ci a malheu scolaire et les informat réserves. Faute de pouvo qu'autorisent des obser on se contentera, dans

Pour ne pas alourdir retenu des expériences q

III

Trois hypothèses son Objectiver les critères — une évaluation plus

(1) Ces professeurs ont u suivantes : mathématiques, tibles d'être chargés du cou
 (2) On traite de l'objectivi directement possible les rés

avec précision des qualités
doit être justifiée.

ne, les professeurs d'un
groupe. C'est pourquoi

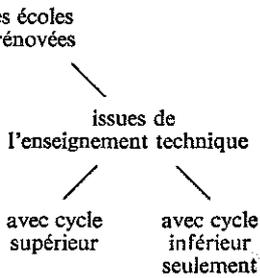
ont on pouvait supposer

ait avant la rénovation :

t que le cycle inférieur
de régents ou de licenciés
(ement).

ans l'école).

en considération dans la
ter schématiquement les



1
.
.
.
.
n

1
.
.
.
.
n

Les trois groupes ainsi formés sont donc équivalents en ce qui concerne les trois critères d'échantillonnage retenus.

2° Grandeur de l'échantillon.

Pour la première expérience, nous avons pris contact avec 700 professeurs de langue maternelle et 1050 professeurs de sciences (1), répartis dans les 132 écoles ayant entrepris la rénovation au moment de l'étude.

Pour ce genre d'entreprise, le pourcentage de réponses favorables a été élevé : 36 % des professeurs de langue maternelle et 33 % des professeurs de sciences.

II. — SECONDE EXPÉRIMENTATION

Sur la base des résultats obtenus lors de la première expérimentation, des remaniements ont été opérés et les règles méthodologiques de construction de l'instrument ont été modifiées.

Les instruments remaniés ont fait l'objet de la deuxième expérimentation. Celle-ci a malheureusement été réalisée tardivement dans l'année scolaire et les informations qu'elle a apportées appellent parfois des réserves. Faute de pouvoir retirer de l'analyse des résultats des certitudes qu'autorisent des observations menées sur un grand nombre de sujets, on se contentera, dans plusieurs cas, de dégager des tendances.

Pour ne pas alourdir encore une présentation déjà pesante, on n'a retenu des expériences que les résultats relatifs aux hypothèses à vérifier.

III. — HYPOTHÈSES À VÉRIFIER

Trois hypothèses sont successivement examinées.

Objectiver les critères d'évaluation permet :

— une évaluation plus objective (2) (hypothèse 1);

(1) Ces professeurs ont une formation de régent ou de licencié dans les matières suivantes : mathématiques, physique, chimie, géographie, biologie. Tous sont susceptibles d'être chargés du cours de travaux scientifiques expérimentaux.

(2) On traite de l'objectivité de l'évaluation avant la validité pour présenter le plus directement possible les résultats expérimentaux.

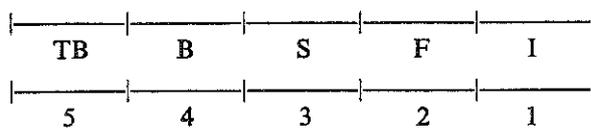
- une évaluation plus valide (hypothèse 2);
- une évaluation mieux utilisable par le professeur et par l'élève (hypothèse 3).

Les effets des échelles descriptives seront appréciés à travers deux types d'informations :

- La *démarche analytique* traduite par les notes relatives aux différents critères; les commentaires des professeurs qui accompagnent la note de synthèse seront également étudiés.
- La *note de synthèse* attribuée au travail.

IV. — DÉPOUILLEMENT DES DONNÉES

Afin de pouvoir analyser statistiquement les données, toutes les appréciations formulées ont été traduites en chiffres :



Après avoir évalué :

- les professeurs des groupes I et III devaient justifier la note globale;
- les professeurs du groupe II devaient, non seulement justifier la note globale, mais aussi chacune des appréciations analytiques.

Les justifications ainsi obtenues ont fait l'objet d'une analyse de contenu.

Pour les justifications des appréciations globales, nous avons d'abord utilisé les catégories choisies dans le dossier : pertinence sémantique, pertinence syntaxique,... en langue maternelle; en sciences, les catégories sont : relation de l'expérience par une description écrite ou un schéma, interprétation...

Les justifications échappant à ces catégories, mais se présentant fréquemment, sont également retenues. Par exemple, on a tenu compte, pour la langue maternelle, de catégories telles que l'orthographe, la ponctuation, la conjugaison, le style,... et, pour les sciences : la qualité

de l'expression écrite, l'expérience,...

Pour les appréciations aspects prévus dans les et « Intégration », pour précis et chronologique justifications proposées négligées.

Enfin, pour chaque les informations suivantes :
 — régent ou licencié;
 — type d'établissement institut technique;
 — localisation de l'école;
 — discipline enseignée;
 — spécialité des professeurs chimie, géographie

Hypothèse 1 : L'é

Comme on l'a déjà comme signe de l'ob par différents juges.

Les professeurs co une performance d'é de tout critère d'éva bablement selon une c

(¹) Bien que l'échelle hypothèses, des techniques d'échelles d'intervalle et de Kerlinger, notamment) de mesure plus rigoureuse, c

de l'expression écrite, de l'observation, la compréhension de l'objet de l'expérience,...

Pour les appréciations analytiques, on a considéré, en priorité, les aspects prévus dans les échelles : « Nombre d'éléments non pertinents » et « Intégration », pour la pertinence sémantique; « Caractère complet, précis et chronologique de la description », en sciences. Les autres justifications proposées par les professeurs n'ont cependant pas été négligées.

Enfin, pour chaque professeur participant à la recherche, on a réuni les informations suivantes :

- régent ou licencié;
- type d'établissement scolaire : athénée et lycée; école moyenne ou institut technique;
- localisation de l'école par province;
- discipline enseignée : français ou sciences;
- spécialité des professeurs de sciences : mathématique, physique, chimie, géographie, biologie.

V. — RÉSULTATS

Hypothèse 1 : L'échelle descriptive permet une évaluation plus objective

Comment vérifier cette hypothèse ?

Comme on l'a déjà souligné à plusieurs reprises, on retient souvent, comme signe de l'objectivité, la concordance des évaluations portées par différents juges.

Les professeurs collaborant à la recherche ont été invités à évaluer une performance d'élève sur une échelle à cinq degrés. En l'absence de tout critère d'évaluation, l'ensemble des notes se distribuera probablement selon une courbe de Gauss (ou selon une ébauche de celle-ci).

(¹) Bien que l'échelle utilisée soit ordinale, on utilisera, pour l'examen des trois hypothèses, des techniques statistiques généralement réservées aux données issues d'échelles d'intervalle et de rapport. Des psychométriciens de grand renom (Guilford, Kerlinger, notamment) ont montré que, moyennant certains efforts pour rendre la mesure plus rigoureuse, cette liberté se justifie par la validité des résultats obtenus.

Par contre, si les juges adoptent les mêmes critères d'évaluation, on peut espérer une concentration des notes sur un même échelon.

Malheureusement, pareille unanimité est rarement observée dans la réalité. L'objectif reste cependant de s'en approcher le plus possible.

L'indice statistique qui rend compte de la plus ou moins grande dispersion des notes attribuées par des juges est l'écart-type (σ). Pour tester l'efficacité des critères d'évaluation, il importe donc de comparer les indices de dispersion des notes analytiques obtenues dans les groupes II et III. A titre indicatif, on comparera également les distributions des notes globales dans les trois groupes.

Mais relever des différences dans la distribution des notes ne suffit pas : encore faut-il déterminer si elles s'expliquent par des fluctuations dues au hasard ou, au contraire, par des variations significatives des conditions dans lesquelles chacun des groupes a évalué.

En outre, les informations fournies par l'analyse des indices de dispersion doivent être complétées par une étude plus approfondie de la répartition des notes. En effet, les juges peuvent s'accorder sur des échelons différents de celui auquel les critères devraient conduire. Pour mieux visualiser cette répartition, les *histogrammes de fréquence* ont été établis pour chaque trait évalué.

Après la comparaison des indices de dispersion pour chaque critère, on comparera la série, le *pattern* des notes analytiques portées par un juge au profil imposé par l'échelle descriptive.

PREMIÈRE EXPÉRIMENTATION (1)

1° CONCORDANCE DES NOTES PARTIELLES, CONSIDÉRÉES ISOLÉMENT.

a) *Comparaison des appréciations analytiques formulées par les groupes II et III.*

Avant d'aborder l'interprétation des résultats, il faut noter que les conditions expérimentales dans lesquelles ils ont été obtenus sont de

(1) On a reporté, en annexe, le détail des analyses statistiques. Dans le présent chapitre, on n'a mentionné que les faits dominants.

nature à diminuer la connaissance du dossier, la procédure ne garantit pas l'évaluation et un respect

Pour la *langue maternelle*

1. *L'échelle portant sur les*
Dans le groupe II, qui évalue de chaque échelon, 61 % (tendance centrale; moyenne des échelons définis), 65 % de (3,7).

2. *Pertinence sémantique.*
Même phénomène que p

3. *Pertinence syntaxique.*
La précision apportée a mais un peu moins nettement l'échelon B, tandis

Les deux autres échelles

1. *Organisation des idées :*
Les notes se concentrent ambiguïtés dans la définit

2. *Vocabulaire.*
Ici, aucun progrès sensiblement les degrés de l'adéquation des échelles et les enseignants moins voulu tenter l'expér

Pour les *sciences*, les résultats maternelle. En gros, le gr

b) *Comparaison des app*
Rappelons que, dans formulée librement, san

En *langue maternelle*, sont nettement plus accusés le groupe II et le groupe

Par contre, en *sciences*

es critères d'évaluation,
ur un même échelon.

ement observée dans la
rocher le plus possible.
plus ou moins grande
est l'écart-type (σ). Pour
porte donc de comparer
enues dans les groupes II
ent les distributions des

ation des notes ne suffit
ent par des fluctuations
iations significatives des
a évalué.

alyse des indices de dis-
plus approfondie de la
vent s'accorder sur des
levraient conduire. Pour
mmes de fréquence ont

ion pour chaque critère,
alytiques portées par un

(1)

ÉRÉES ISOLÉMENT.

mulées par les groupes II

ts, il faut noter que les
nt été obtenus sont de

statistiques. Dans le présent

nature à diminuer la concordance attendue. Les professeurs ont pris connaissance du dossier par une seule lecture individuelle. Une telle procédure ne garantit pas une pénétration approfondie des critères d'évaluation et un respect strict des consignes d'utilisation.

Pour la *langue maternelle*, trois échelles se révèlent immédiatement efficaces :

1. *L'échelle portant sur les structures grammaticales.*

Dans le groupe II, qui évalue sans pouvoir s'appuyer sur une définition de chaque échelon, 61 % des notes se concentrent sur le milieu de l'échelle (tendance centrale; moyenne proche de 3). Par contre, dans le groupe III (échelons définis), 65 % des appréciations se situent sur l'échelon B (moyenne 3,7).

2. *Pertinence sémantique.*

Même phénomène que pour l'échelle précédente.

3. *Pertinence syntaxique.*

La précision apportée aux critères joue de nouveau dans le même sens, mais un peu moins nettement. La majorité des professeurs du groupe II choisissent l'échelon B, tandis que, dans le groupe III, elle se porte sur TB.

Les deux autres échelles déçoivent.

1. *Organisation des idées :*

Les notes se concentrent mieux dans le groupe III, mais il apparaît que des ambiguïtés dans la définition des degrés ont réduit la netteté des effets.

2. *Vocabulaire.*

Ici, aucun progrès sensible n'est enregistré. Il est difficile de définir clairement les degrés de l'adéquation et de la richesse du vocabulaire. Les constructeurs des échelles et les enseignants qui les ont aidés le savaient. Ils ont néanmoins voulu tenter l'expérience : elle a échoué.

Pour les *sciences*, les résultats sont moins favorables que pour la *langue maternelle*. En gros, le groupe II tend à mieux évaluer que le groupe III.

b) *Comparaison des appréciations globales dans les trois groupes.*

Rappelons que, dans les trois groupes, l'appréciation globale a été formulée librement, sans support critériel.

En *langue maternelle*, les différences de dispersion dans les notes sont nettement plus accusées entre le groupe I et les deux autres, qu'entre le groupe II et le groupe III.

Par contre, en *sciences*, le meilleur accord est observé dans le groupe I.

RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

Les échelles ont été remaniées en fonction des observations qui précèdent.

a) *Comparaison des appréciations analytiques formulées par les groupes II et III.*

En *langue maternelle*, la concordance entre les juges augmente fortement dans le groupe III, sauf pour l'échelle de vocabulaire. Pour cette dernière, le critère objectif le plus aisément accessible serait la comparaison avec un vocabulaire de base.

En *sciences*, deux échelles donnent maintenant entière satisfaction : « Relation écrite de l'expérience » et « Description de l'expérience par un schéma ».

Pour les deux autres échelles (« Interprétation » et « Invention »), le progrès est moindre. Ce phénomène semble toutefois accidentel : dans le travail d'élève à juger, les réponses ne sont pas fournies dans l'ordre des questions; l'interview de plusieurs professeurs révèle que des confusions se sont ainsi créées.

b) *Comparaison des appréciations globales.*

Les résultats de la première expérimentation se confirment.

2^o ÉTUDE DE LA CONCORDANCE ENTRE LES PROFILS DE NOTES.

Jusqu'à présent, chaque évaluation a été considérée isolément. Il est aussi intéressant de voir si, devant évaluer analytiquement à plusieurs égards, les professeurs aboutissent à des profils de notes superposables.

Si l'on envisage tous les critères de l'évaluation, les résultats peuvent sembler imparfaits. En fait, ce phénomène était attendu : on imagine, en effet, quel degré de perfection il faudrait atteindre dans la construction de l'instrument pour que toutes les notes s'harmonisent parfaitement pour tous les aspects retenus.

Il est donc remarquable de constater qu'en négligeant l'échelle « Vocabulaire » en langue maternelle, la deuxième mise au point entraîne des progrès importants.

Langue maternelle ...

Sciences

Pourcentages d

Premières conclusions.

Si l'objectif est d'analyser les appréciations globales que des juges proposent, les analyses précédentes fournissent la définition de la concordance des degrés, pourcentage moindre.

Pourtant, si l'on veut comparer les appréciations analytiques des échelons reste primaires globale et de façon analytique, l'importance relative de l'évaluation.

3^o ÉTUDE DE L'INFLUENCE

a) *Les échelles descriptives centrale ?*

1. *Evaluation analytique.*

En langue maternelle, la concordance de chacun des échelons (grâce à la concordance centrale (sauf pour l'échelle de vocabulaire))

En sciences, les résultats sont caractéristiques du travail

des observations qui

formulées par les groupes II

juges augmente fortement
 ire. Pour cette dernière, le
 omparaison avec un voca-

tière satisfaction : « Rela-
 périece par un schéma ».

« Invention »), le progrès
 cidental : dans le travail
 ans l'ordre des questions;
 confusions se sont ainsi

se confirment.

LS DE NOTES.

considérée isolément. Il est
 analytiquement à plusieurs
 de notes superposables.

on, les résultats peuvent
 t attendu : on imagine,
 eindre dans la construc-
 s'harmonisent parfaite-

en négligeant l'échelle
 e mise au point entraîne

	1 ^{re} expérimentation		2 ^e expérimentation
	Groupe II	Groupe III	Groupe III
Langue maternelle	34	60	63
Sciences	24	35	68

Pourcentages d'accords avec le profil idéal (nombres arrondis)

Premières conclusions.

Si l'objectif est d'augmenter la concordance des appréciations globales que des juges portent sur une performance complexe, il ressort des analyses précédentes que la démarche la plus économique est de leur fournir la définition des critères à prendre en considération. La spécification des degrés, pour chacun de ces critères, apporte un bénéfice moindre.

Pourtant, si l'on veut rendre plus objectives et plus nuancées les appréciations analytiques des performances complexes, la description des échelons reste primordiale. L'idéal est de noter à la fois de façon globale et de façon analytique; l'élève reçoit ainsi des informations sur l'importance relative que le professeur accorde aux divers aspects de l'évaluation.

3^o ÉTUDE DE L'INFLUENCE DES FACTEURS EXTERNES.

a) *Les échelles descriptives aident-elles à lutter contre la tendance centrale ?*

1. *Evaluation analytique.*

En langue maternelle, les histogrammes montrent que la description de chacun des échelons (groupe III) permet de lutter efficacement contre la tendance centrale (sauf pour le vocabulaire).

En sciences, les résultats — bien que plus facilement acquis en raison des caractéristiques du travail de l'élève — sont également satisfaisants.

2. Notes globales.

En français comme en sciences, la moyenne des notes globales s'écarte de plus en plus de la note centrale à mesure que l'on passe du groupe I au groupe II et du groupe II au groupe III. L'expérience est donc positive.

3. Profils des notes.

En langue maternelle, le groupe III s'écarte le plus de la note centrale.

b) Les échelles descriptives aident-elles à lutter contre l'effet de contamination ?

ANALYSE DES CORRÉLATIONS.

— Langue maternelle.

L'analyse des corrélations entre notes analytiques conduit aux observations suivantes :

1. Dans le groupe II, les professeurs ne dissocient en général pas bien « l'organisation des idées » de tous les autres critères. On constate aussi une forte liaison entre le « vocabulaire » et la « pertinence sémantique ». Le critère « structure grammaticale » se révèle, par contre, pratiquement indépendant de tous les autres.
2. Dans le groupe III, aucun critère n'est fortement corrélé avec tous les autres. L'effort d'objectivation des critères a donc porté. En particulier, la « pertinence sémantique » réussit maintenant à jouer le rôle particulier qu'on lui assignait.
« Pertinence syntaxique » et « structures grammaticales » restent un peu liées (.28), ce qui paraît dans la normale des choses, car les deux variables se complètent.

— Sciences.

Dans les deux groupes, on trouve une corrélation significative entre la « relation écrite de l'expérience » et la « relation de l'expérience par schémas ». Il ne semble pas que l'on ait affaire à un effet de contamination; les deux variables sont réellement parentes. « L'interprétation » et l'« invention » sont, de leur côté, liées. Ce phénomène semble lui aussi normal, car il s'agit de processus intellectuels supérieurs dont on sait la saturation en facteur d'intelligence générale.

ANALYSE FACTORIELLE.

Cette technique permet de déterminer la parenté entre plus de deux variables. Les résultats suivants doivent être accueillis avec des réserves puisque les appréciations n'ont porté que sur un seul travail d'élève. L'allure des résultats peut autant dépendre des caractéristiques particulières du travail que des critères de correction.

— Langue maternelle.

Pour le groupe II, une saturation des trois facteurs révèle déjà émise : pour évaluer d'aspects qu'ils ont déjà é

On observe, en outre, une sémantique et le vocabulaire l'organisation des idées.

La situation se clarifie d'abord la présence d'un notes à laquelle l'application ambitionne de conduire.

Les deux dimensions de « organisation des idées » spécifique.

— Sciences.

La clarification est ne deux facteurs apparaissent (description écrite et description cognitives supérieures (int

40 ÉTUDE DE L'INFLUENCE

Les deux principales sont dans l'intervention de (nous venons d'en faire de critères, variables sel

Faut-il encore rappeler échelles, on cherche à professeurs adoptent. On critères d'évaluation ont possible. Les informations les groupes II et III.

Pour évaluer, les juges ceux qu'on leur prop

Seul un entretien app par l'expérience permet

— *Langue maternelle.*

Pour le groupe II, une constatation domine : « L'organisation des idées » sature les trois facteurs révélés par l'analyse. Ainsi se confirme une hypothèse déjà émise : pour évaluer selon ce critère, les professeurs tiennent compte d'aspects qu'ils ont déjà évalués ailleurs.

On observe, en outre, une parenté factorielle entre, d'une part, la pertinence sémantique et le vocabulaire et, d'autre part, la pertinence syntaxique et l'organisation des idées.

La situation se clarifie considérablement dans le groupe III. On observe d'abord la présence d'un facteur général qui reflète bien l'homogénéité des notes à laquelle l'application correcte des critères d'évaluation proposés ambitionne de conduire.

Les deux dimensions relatives au contenu (« pertinence sémantique » et « organisation des idées ») ressortent, en outre, comme un autre facteur spécifique.

— *Sciences.*

La clarification est nette : pour le groupe II comme pour le groupe III, deux facteurs apparaissent : le premier regroupe les deux critères descriptifs (description écrite et description par schémas) et le second les deux opérations cognitives supérieures (interprétation et invention).

4^e ÉTUDE DE L'INFLUENCE DES FACTEURS INTERNES.

Les deux principales sources de discordance résident, nous l'avons vu, dans l'intervention de facteurs étrangers à la démarche d'évaluation (nous venons d'en faire l'analyse) ou peuvent s'expliquer par un choix de critères, variables selon les juges. Il s'agit alors de facteurs internes.

Faut-il encore rappeler qu'en définissant l'objet et les degrés des échelles, on cherche à homogénéiser les critères d'évaluation que les professeurs adoptent. On sait aussi que, dans un premier temps, différents critères d'évaluation ont été choisis et définis aussi rigoureusement que possible. Les informations données à ce propos ont été les mêmes pour les groupes II et III.

Pour évaluer, les juges n'ont-ils pas fait intervenir d'autres critères que ceux qu'on leur proposait ?

Seul un entretien approfondi avec chacun des professeurs concernés par l'expérience permettrait de répondre complètement à cette question.

Faute de pouvoir entreprendre une telle étude, nous avons cherché à connaître, par analyse statistique, la part prise par l'ensemble des notes analytiques dans l'attribution de la note de synthèse ⁽¹⁾.

Les résultats ne sont guère probants, surtout en sciences. Ils semblent indiquer qu'au moment de l'appréciation globale, les professeurs se sentent libérés des contraintes analytiques et font intervenir des critères personnels, plus ou moins éloignés de ceux qu'on leur avait proposés avant.

Hypothèse 2 : L'échelle descriptive permet une évaluation plus valide

Comment vérifier cette hypothèse ?

Pour construire les échelles, nous avons d'abord consulté des spécialistes de l'enseignement des branches concernées, sur les différents aspects à considérer dans l'évaluation. Mais, même si tous les aspects importants ont été retenus, il n'est pas acquis d'avance que les utilisateurs de l'instrument tiendront compte de tout au moment de formuler leur note globale. En particulier, on peut poser les questions suivantes :

- La note globale est-elle significativement corrélée avec chacune des appréciations analytiques (analyse des corrélations simples) ?
- Quelle influence respective exercent les notes analytiques sur la note globale (analyse de régression multiple, analyse par segmentation, analyse factorielle) ?

1° CORRÉLATION ENTRE LA NOTE GLOBALE ET CHACUNE DES APPRÉCIATIONS ANALYTIQUES.

— *Langue maternelle :*

Par rapport au groupe II, dans lequel certains critères (pertinence sémantique, vocabulaire, organisation des idées) sont corrélés avec la note globale, au détriment de la pertinence sémantique et des structures grammaticales, tous les critères utilisés dans le groupe III sont significativement corrélés avec l'appréciation globale.

(1) Cette importance relative s'exprime par le pourcentage de variance expliquée.

— *Sciences :*

Les coefficients de corrélation des appréciations globales sont faibles, surtout pour la description et l'interprétation similaire à l'écrit. Le pourcentage de variance expliquée, dans les deux disciplines, est de 10 %.

2° INFLUENCE RESPECTIVE DES NOTES ANALYTIQUES SUR LA NOTE GLOBALE

Plusieurs traitements statistiques ont été utilisés pour répondre à cette question. Nous ne présentons ici que les conclusions permises par l'analyse de régression multiple.

Analyse de régression.

Dans les deux disciplines, la note globale est plus corrélée avec le groupe II que dans le groupe III. La dépendance des notes dans les deux disciplines est plus forte dans le groupe II que dans le groupe III.

Analyse factorielle.

Dans le groupe III, la note globale est très saturée dans toutes les dimensions de l'échelle. L'équilibre des notes analytiques est plus satisfaisant que dans le groupe II.

En sciences, les appréciations globales sont plus corrélées vis-à-vis de la note globale que dans le groupe II. L'appréciation appartenant au même facteur est plus corrélée (écrite et par schémas) dans le groupe II que dans le groupe III. Par contre, la note globale est plus corrélée avec l'appréciation « invention » et « invention » que dans le groupe II.

Analyse par segmentation.

En langue maternelle, la note globale est plus corrélée dans les deux populations que dans le groupe II. Les appréciations sont corrélées différemment. En français, la note globale est plus corrélée avec l'appréciation « écriture » de plus, un apport plus important est fait à la note globale.

En sciences aussi, la note globale est plus corrélée basent leur appréciation globale sur les notes analytiques. Les collègues du groupe III ont une appréciation globale plus corrélée avec les notes analytiques et, notamment, avec l'appréciation « invention ».

le, nous avons cherché
prise par l'ensemble des
de synthèse (1).

en sciences. Ils semblent
male, les professeurs se
nt intervenir des critères
on leur avait proposés

évaluation plus valide

èse ?

ord consulté des spécia-
nées, sur les différents
même si tous les aspects
avance que les utilisateurs
moment de formuler leur
questions suivantes :

rrélée avec chacune des
ations simples) ?

analytiques sur la note
analyse par segmentation,

CHACUNE DES APPRÉCIA-

es (pertinence sémantique,
e la note globale, au détri-
s grammaticales, tous les
ent corrélés avec l'appré-

atage de variance expliquée.

— *Sciences :*

Les coefficients de corrélation entre les appréciations analytiques et les appréciations globales sont plus élevés dans le groupe II que dans le groupe III, surtout pour la description écrite. Ce résultat inattendu appelle sans doute une interprétation similaire à celle que l'on a proposée à propos du pourcentage de variance expliquée, dans les deux groupes, pour les notes analytiques.

2° INFLUENCE RESPECTIVE DES NOTES ANALYTIQUES SUR LA NOTE GLOBALE.

Plusieurs traitements statistiques ont été réalisés pour répondre à cette question. Nous nous contenterons de dégager ici l'essentiel des conclusions permises par ces analyses.

Analyse de régression.

Dans les deux disciplines, la part de variance jointe, plus importante dans le groupe II que dans le groupe III, plaide en faveur d'une plus grande indépendance des notes dans le groupe expérimental.

Analyse factorielle.

Dans le groupe III, la présence de la note globale dans le premier facteur, très saturé dans toutes les variables, souligne une nouvelle fois la contribution équilibrée des notes analytiques à la note globale.

En sciences, les appréciations analytiques jouent un rôle très différent vis-à-vis de la note globale. Dans le groupe de contrôle, les deux variables appartenant au même facteur que la note globale sont relatives à la description (écrite et par schémas) des phénomènes observés. Dans le groupe expérimental, par contre, la note globale procède du même facteur que les variables « interprétation » et « invention ».

Analyse par segmentation.

En langue maternelle, la pertinence sémantique joue un rôle prépondérant dans les deux populations. Au-delà de ce critère, les deux groupes se comportent différemment. En effet, les autres variables jouent un rôle nettement plus significatif dans le groupe III que dans le groupe II. Ceci confirme, une fois de plus, un apport plus équilibré des appréciations analytiques à la note globale.

En sciences aussi, alors que bon nombre de professeurs du groupe II ne basent leur appréciation globale que sur l'interprétation et la description, leurs collègues du groupe III font plus largement appel à toutes les appréciations analytiques et, notamment, à l'invention.

Hypothèse 3 : L'échelle descriptive permet une évaluation mieux utilisable par le professeur et par l'élève

Comment vérifier cette hypothèse ?

La façon dont les professeurs justifient leur note globale permet le mieux de vérifier cette hypothèse.

L'évaluation formative suppose que les objectifs poursuivis soient clairement définis et communiqués aux élèves. Au terme d'un travail, il importe que la justification de la note attribuée soit en rapport direct avec ces objectifs. De plus, la formulation précise et, autant que possible, positive de cette justification doit aider l'élève pour la suite de son apprentissage.

1° LE NOMBRE DE JUSTIFICATIONS DIFFÈRE-T-IL D'UN GROUPE À L'AUTRE ?

Elles sont les plus nombreuses dans le groupe I où elles constituent le seul moyen laissé aux professeurs pour expliquer leur note. Dans les autres groupes, elles sont d'autant moins nombreuses que les professeurs ont déjà eu l'occasion d'expliquer la note globale par une démarche analytique préalable. Ce phénomène est accentué dans le groupe II où les professeurs avaient été invités à justifier chacune de leurs appréciations analytiques.

2° LES PROFESSEURS UTILISENT-ILS DES JUSTIFICATIONS EN RAPPORT AVEC LES OBJECTIFS DU TRAVAIL PROPOSÉ ?

En langue maternelle, la proportion de justifications principales par rapport aux justifications secondaires est nettement plus importante dans le groupe III. Cette répartition est surtout due au poids que prennent l'orthographe et la ponctuation dans le groupe I. De plus, on constate un meilleur équilibre des différentes justifications principales dans le groupe III.

Les questions posées à l'élève, pour le travail de sciences, indiquent plus clairement les critères de l'évaluation. Cette particularité explique que les professeurs du groupe I appuient fréquemment leurs notes globales par des justifications principales. Cependant, les professeurs du groupe III accordent moins d'importance à l'invention : ce comporte-

ment semble s'accorder pédagogiquement.

3° LES COMMENTAIRES DE FACILEMENT RECONNAISSABLES

Différentes analyses, parmi les justifications qui peu sont suffisamment précises pour la suite de son apprentissage à des détails. Le pourcentage est relativement élevé.

Dans ces conditions, le commentaire présenté à l'élève sur les échelles descriptives analytiques se justifie, attribution d'un degré dans une double référence. La description de chaque degré de l'échelle d'une habileté et permet le rapport à l'objectif à atteindre en fonction des possibilités de l'élève. L'évaluation sa référence à la justification de l'appréciation explicite sur la direction à progresser et, éventuellement,

4° LES PROFESSEURS RECHERCHENT-ILS DES ASPECTS POSITIFS OU NÉGATIFS ?

Pour les deux disciplines, il est nettement plus important dans les deux groupes, dans lesquels les aspects négatifs.

En outre, les professeurs de langue maternelle, sur le caractère particulier du travail. Dans le groupe III, la note relevée en positif par ce

évaluation mieux utilisable

à l'élève ?

une note globale permet le

objectifs poursuivis soient
Au terme d'un travail,
soit en rapport direct
et, autant que possible,
pour la suite de son

UN GROUPE À L'AUTRE ?

de I où elles constituent
leur note. Dans les
classes où les professeurs
globale par une démarche
dans le groupe II où
aucune de leurs apprécia-

JUSTIFICATIONS EN RAPPORT AVEC

justifications principales par
est nettement plus importante
due au poids que pren-
le groupe I. De plus, on consi-
dérations principales dans le

de sciences, indiquent
la particularité explique
surtout leurs notes glo-
bales, les professeurs du
groupe I ont une justification : ce comporte-

ment semble s'accorder davantage aux exigences requises à ce niveau pédagogique.

3° LES COMMENTAIRES DESTINÉS À L'ÉLÈVE CONCERNENT-ILS DES CRITÈRES FACILEMENT RECONNAISSABLES PAR CE DERNIER ?

Différentes analyses, faites pour la langue maternelle, montrent que parmi les justifications que l'on peut classer dans une catégorie définie, peu sont suffisamment spécifiques pour que l'élève puisse en tirer parti pour la suite de son apprentissage. Certaines s'attachent uniquement à des détails. Le pourcentage des justifications inclassables est relativement élevé.

Dans ces conditions, quel parti l'élève peut-il tirer d'un seul commentaire présenté à l'appui d'une note globale ? Ici l'avantage des échelles descriptives apparaît nettement : chacune des appréciations analytiques se justifie, automatiquement et sans ambiguïté, par l'attribution d'un degré dans l'échelle. Or les instruments permettent une double référence. La description des exigences requises pour l'attribution de chaque degré de l'échelle fixe les points critiques du développement d'une habileté et permet donc à l'élève de situer sa performance par rapport à l'objectif à atteindre. De plus, l'adaptation des niveaux d'exigence aux possibilités d'élèves d'un âge pédagogique déterminé, donne à l'évaluation sa référence normative. Si ces précisions sont apportées, la justification de l'appréciation globale peut se résumer en une courte explication sur la direction principale des apprentissages à réaliser pour progresser et, éventuellement, en un commentaire de caractère affectif.

4° LES PROFESSEURS RELÈVENT-ILS, DANS LEURS COMMENTAIRES, LES ASPECTS POSITIFS OU NÉGATIFS DU TRAVAIL ?

Pour les deux disciplines, la proportion des justifications positives est nettement plus importante dans le groupe III que dans les deux autres groupes, dans lesquels les justifications secondaires sont souvent négatives.

En outre, les professeurs du groupe III s'accordent mieux, en langue maternelle, sur le caractère positif ou négatif d'une caractéristique particulière du travail. Dans le groupe I, la même caractéristique est souvent relevée en positif par certains et en négatif par d'autres.

EN RÉSUMÉ

Au terme de cette étude, quelques observations s'imposent avec force. Il est utopique de vouloir éliminer toute discordance dans l'évaluation des performances complexes. Néanmoins, il est possible de réduire considérablement les divergences entre correcteurs.

L'instrument construit ne porte certainement pas sur les seuls critères pertinents pour apprécier les exercices choisis. Il se veut, avant tout, occasion de réflexion; chacun pourra éventuellement amender, selon ses besoins, ses expériences personnelles, ses critiques.

En dernière analyse, la méthode de construction proposée importe plus que le cas particulier choisi pour la démonstration. On peut raisonnablement penser que, correctement appliquée, la méthodologie proposée apporte une garantie de validité des échelles d'évaluation.

En général, un professeur isolé dispose rarement du temps nécessaire pour construire des échelles descriptives opérationnelles et pour les valider expérimentalement. Des équipes de travail doivent donc se constituer partout où cela est possible.

De cette façon, et à condition que l'information circule correctement, une gamme de plus en plus large d'instruments d'évaluation verrait le jour et serait mise à la disposition de tous.

Faut-il dire que les instruments complexes que sont les échelles d'évaluation descriptives ne sont pas faits pour être utilisés à toute heure du jour... Mais employés à intervalles réguliers, ils peuvent fournir des points de repère objectifs et valides susceptibles de compléter les observations faites, par ailleurs, par les professeurs.

CONCLUSION

Un peu paradoxalement, nos principales conclusions figurent au terme... de l'introduction du présent ouvrage : possibilité d'augmenter la rigueur de l'évaluation des performances complexes et, par là, la concordance des appréciations; nécessité de modifier certaines façons d'évaluer, parfois pratiquées depuis longtemps; lourdeur du travail de construction d'instruments de mesure quelque peu perfectionnés;

besoin d'une collaborat
et les services de recher

On l'a déjà dit aussi,
tives à leur constructio
réalité est beaucoup plu
bien connu de nous to
dédale technique des m
manier !

Prétendre que tout es
l'évaluation rénovés al
faux. Le travail des en
de plus en plus comple
et continue est intelle
Moins que jamais, les s
leur place à l'école. Ce
plus l'éducation gagne e
tâche devient difficile. I

besoin d'une collaboration de plus en plus étroite entre les enseignants et les services de recherche.

On l'a déjà dit aussi, les échelles proposées et les considérations relatives à leur construction sont, en apparence, compliquées. En fait, la réalité est beaucoup plus simple que sa description verbale, phénomène bien connu de nous tous qui, de temps en temps, nous perdons dans le dédale technique des modes d'emploi d'appareils pourtant si faciles à manier !

Prétendre que tout est simple, que la pratique de l'enseignement et de l'évaluation rénovés allège la tâche des professeurs serait cependant faux. Le travail des enseignants qui veulent bien faire leur métier est de plus en plus complexe et dur; en particulier, l'évaluation formative et continue est intellectuellement et matériellement très exigeante. Moins que jamais, les solutions de facilité et le confort intellectuel n'ont leur place à l'école. Ce n'est pas le moindre paradoxe de notre métier : plus l'éducation gagne en liberté, en spontanéité, en créativité, plus notre tâche devient difficile. La société se doit de le reconnaître.

ANNEXES

ANNEXE A
ÉCHELLES D'ÉVALUATION

LANGUE MATERNELLE

Enseignement secondaire rénové : 1^{er} cycle

ÉCHELLES D'ÉVALUATION

RÉDACTION :

COMPLÈTEMENT D'UN RÉCIT MUTILÉ

Ce dossier est dû à la collaboration d'une commission de professeurs composée de :

Mmes DENOISEUX, E.M.E. Flémalle;
GILLIS, E.M.E. Liège ;
HAUTOT, A.R. Bouillon;
KAUFFMANN, E.M.E. Arlon;
LAURENT, E.M.E. Arlon;
MELIGNON, A.R. Liège II;
RAEMACKERS, E.M.E. Waremme.

MM. REMACLE, E.M.E. Dolhain-Limbourg;
RENSON, A.R. Laroche;
THUMILAIRE, I.E.T.E. Libramont.

Chercheurs : J. PAQUAY-BECKERS et R. DE BAL.

L'évaluation fait partie intégrante du processus d'enseignement dont elle est une composante essentielle. Aussi importe-t-il de garantir la *validité* des instruments de mesure utilisés. Un instrument valide est représentatif de la matière dont on souhaite vérifier l'apprentissage : les questions explorent à la fois tous les contenus et tous les comportements applicables à ces contenus. Une telle procédure nécessite des études théoriques qui recensent les éléments fondamentaux de l'acquisition d'une notion. L'évaluation permet alors d'informer chaque élève de son progrès vers un objectif déterminé. En outre, la performance de l'élève peut être comparée à celles d'une population de référence (étalonnage normatif). Nous sommes encore loin de disposer des études et des instruments nécessaires...

Pour remplir son rôle vis-à-vis de l'élève, l'évaluation doit être aussi *objective* et *nuancée* que possible. Le degré de concordance entre différents juges est un indice important de l'objectivité d'une note. Or de nombreuses études docimologiques ont montré, pratiquement pour toutes les disciplines, combien les juges s'accordent peu sur la note à attribuer à une performance donnée. De plus, l'élève ne peut tirer un profit réel de l'évaluation que si les informations qui l'accompagnent sont analytiques, complètes et précises. Pour que le professeur puisse apporter un tel *feedback* à l'élève, la démarche qui préside à l'attribution de la note doit, elle aussi, être analytique.

Les difficultés de l'évaluation sont d'autant plus grandes que la performance sollicitée de l'élève nécessite la mise en œuvre de connaissances et d'aptitudes nombreuses. Les échelles descriptives présentées ici ont été construites pour l'évaluation de performances complexes. Des professeurs ont mis leur expérience en commun pour choisir des critères d'appréciation aussi sûrs, aussi valides que possible. Ces critères ont été appliqués à de nombreux travaux.

Dans l'espoir d'assurer l'homogénéité des notes, des consignes de correction détaillées sont proposées (et justifiées).

L'évaluation à l'aide d'échelles descriptives fournit deux types d'informations à l'élève :

- Il est informé des points forts et des lacunes de son travail.
- Il dispose de points de repère pour situer sa performance par rapport à ses condisciples.

En raison de son caractère analytique, l'évaluation à l'aide d'échelles descriptives prend beaucoup de temps : elle ne pourrait pas être pratiquée quotidiennement dans une classe. Elle devrait être réservée pour certains moments cruciaux de l'apprentissage, ou, plus simplement, pour des vérifications périodiques. Au lieu de se servir des batteries complètes qui lui sont proposées, le professeur a aussi toujours le loisir de n'utiliser qu'une ou deux des échelles, selon les besoins.

On peut espérer que, petit à petit, se créeront des banques d'instruments d'évaluation auxquelles les enseignants pourront faire appel à leur convenance et à celle de leurs élèves.

Notes pour l'évaluation des travaux.

1. Les critères d'évaluation proposés dans les échelles ci-après ont été définis et testés par la commission de professeurs, au cours de nombreuses séances de travail. Il se peut que ces critères ne soient pas ceux en fonction desquels vous jugez habituellement. Pour essayer l'instrument, il importe cependant que vous vous conformiez strictement aux consignes de correction proposées.

En vue de modifications ultérieures, vos remarques seront précieuses.

2. Pour la commodité et l'efficacité de vos élèves selon la procédure suivante.
3. En guise d'entraînement, un t

d'enseignement dont elle est une
tantir la *validité* des instruments
ésentatif de la matière dont on
orent à la fois tous les contenus
us. Une telle procédure nécessite
fondamentaux de l'acquisition
er chaque élève de son progrès
ce de l'élève peut être comparée
normatif). Nous sommes encore
aires...

luation doit être aussi *objective*
ntre différents juges est un indice
euses études docimologiques ont
ombien les juges s'accordent peu
De plus, l'élève ne peut tirer un
l'accompagnent sont analytiques,
apporter un tel *feedback* à l'élève,
oit, elle aussi, être analytique.

us grandes que la performance
onnaissances et d'aptitudes nom-
été construites pour l'évaluation
nis leur expérience en commun
aussi valides que possible. Ces

es, des consignes de correction

nit deux types d'informations à

on travail,

erformance par rapport à ses

n à l'aide d'échelles descriptives
pratiquée quotidiennement dans
moments cruciaux de l'apprentis-
iodiques. Au lieu de se servir des
seur a aussi toujours le loisir de
ins.

es banques d'instruments d'éva-
pel à leur convenance et à celle

s ci-après ont été définis et testés
nombreuses séances de travail.
nction desquels vous jugez habi-
ependant que vous vous con-
proposées.

En vue de modifications ultérieures des critères, vos remarques et vos critiques seront précieuses.

2. Pour la commodité et l'efficacité de l'évaluation, il est conseillé d'évaluer tous les travaux de vos élèves selon la première échelle, puis de procéder de même pour les suivantes.
3. En guise d'entraînement, un travail corrigé est proposé en annexe.

A. — Description de l'exercice choisi

Au lieu de recourir à une rédaction du genre habituel, on a proposé un récit mutilé en son milieu. Des élèves de première année de l'enseignement secondaire rénové ont été invités à inventer la partie manquante après une lecture individuelle du résumé et des paragraphes d'encadrement.

B. — Justification du choix de cet exercice

Cet exercice demande une performance complexe aux élèves. Son évaluation ne peut se limiter à un simple décompte de points par faute commise; elle nécessite la mise au point d'un instrument plus adéquat.

La préférence a été accordée à ce genre d'exercice plutôt qu'à la rédaction habituelle parce que les contraintes imposées par le résumé et les paragraphes d'encadrement fournissent des points de repère objectifs pour l'évaluation.

C. — Choix des critères d'évaluation

Les deux premiers critères retenus : la pertinence sémantique et la pertinence syntaxique, sont particuliers à ce genre d'exercice. Les trois critères suivants : le vocabulaire, les structures grammaticales et l'organisation des idées dans le paragraphe 2, revêtent un caractère plus général.

L'évaluation selon ces cinq critères ne sera évidemment pas exhaustive. Elle portera néanmoins sur les aspects que plusieurs recherches factorielles font le plus nettement ressortir.

En particulier, on n'a pas construit d'échelle spéciale pour l'appréciation de l'atmosphère, du ton du récit. Il est vraisemblable qu'en notant sur les échelles proposées, les professeurs en auront déjà tenu compte.

L'évaluation des fautes d'orthographe et de conjugaison n'est pas prévue, l'exercice proposé n'étant pas spécialement destiné à mesurer un apprentissage dans ces domaines.

D. — Organisation des dossiers

Pour chacun des critères, on trouvera successivement :

- LA DÉFINITION DU CRITÈRE.
- LA DÉFINITION DES ASPECTS ou DES CARACTÉRISTIQUES qui interviennent dans le critère.
Tous les aspects retenus se voient réserver une importance égale dans l'évaluation.
- LA DÉTERMINATION DES NIVEAUX D'EXIGENCE.
Les exigences, pour chaque degré d'une échelle, ne sont pas arbitraires : elles ont été établies après une analyse approfondie d'un nombre important de travaux d'élèves de première année de l'enseignement secondaire rénové.
- LA DÉMARCHE DE L'ÉVALUATION.
- LA PRÉSENTATION DE L'ÉCHELLE.
Cette présentation n'apporte aucune information complémentaire par rapport à la démarche. Toutefois, elle offre l'avantage de donner une description hiérarchique de la performance.

RÉSUMÉ DES CHAPITRES PRÉCÉDENTS
vient d'arriver à Piriac, en Bretagne, pendant ses vacances chez son oncle Vincent, vieillard qui habite un château isolé, se livre à sa passion :

Il avait été convenu que Georges devait aller accueillir Luc à sa descente de train, à Piriac, distant de près de vingt kilomètres de son oncle. Étonné, tente de téléphoner à son oncle, mais ne trouve à Piriac d'où il gagne le château d'arriver. Luc a beau fureter dans les pièces, personne ne se manifeste. Anna semble s'être volatilisée. Luc poursuit ses recherches dans le

§ 1. — En descendant les premiers sous-sol, Luc sentit son cœur battre. Il s'était aventuré dans cette partie de la cave où il avait eu la tentation de pousser la porte. Ça paraissait chaque après-midi à qu'il se souvienne de l'avertissement du vieillard dans le grenier dont le plancher vibrerait la cave de l'aile gauche où se trouvaient généralement si douce, lui avait permis de penser que jamais il n'avait, malgré ses souvenirs se pressaient dans sa tête vers le sous-sol interdit. « Si au moment où, lorsque, l'étroit passage obliquait vers l'obscurité la plus complète. Il se tenait juste devant lui, une faible lueur qui murmurait-il enfin devant une porte qui filtrait un rayon de lumière crue. Il n'osait pousser, ni même toucher le bois. Un bref instant, ébloui par la vive lumière, il ne discernait rien. Mais, quand ses yeux se levèrent, une brève exclamation à la vue du spectacle.

§ 2. — Paragraphe à rédiger par l'élève

§ 3. — « Mais non, Monsieur, j'ai exploré le moindre recoin; il n'y avait rien. J'ai commencé par ouvrir la petite fenêtre portable. Et c'est en traversant la

(1) Ce texte a été rédigé par M. de l'Institut de Psychologie et de

RÉSUMÉ DES CHAPITRES PRÉCÉDENTS : *Luc, un jeune étudiant belge de quinze ans, vient d'arriver à Piriac, en Bretagne, pour passer, comme chaque été, un mois de vacances chez son oncle Vincent, vieux célibataire un peu original, qui, dans un antique château isolé, se livre à sa passion : la chimie.*

Il avait été convenu que Georges, le fidèle serviteur de l'oncle Vincent, viendrait accueillir Luc à sa descente de train, pour le conduire en voiture de la gare au château, distant de près de vingt kilomètres. Mais personne n'est au rendez-vous. Luc, un peu étonné, tente de téléphoner à son oncle : personne ne répond. Il se rend alors en bus à Piriac d'où il gagne le château à pied. Là-bas, autre sujet d'étonnement : bien qu'il ait trouvé la porte d'entrée largement ouverte, la demeure est apparemment déserte. Luc a beau fureter dans les pièces du rez-de-chaussée et du premier étage, il a beau appeler partout, personne ne se manifeste. L'oncle Vincent, le vieux Georges, la cuisinière Anna semblent s'être volatilisés. Intrigué par ce silence inexplicable, Luc décide de poursuivre ses recherches dans les sous-sols du bâtiment...).

§ 1. — En descendant les premières marches de l'étroit escalier qui conduisait au sous-sol, Luc sentit son cœur battre plus rapidement. Jamais, jusqu'à présent, il ne s'était aventuré dans cette partie du château. Vingt fois, lors de ses séjours à Piriac, il avait eu la tentation de pousser la petite porte sculptée par laquelle son oncle disparaissait chaque après-midi à quinze heures précises. Mais, chaque fois, il s'était souvenu de l'avertissement du vieil homme : « Tu peux te promener partout, sauf dans le grenier dont le plancher vermoulu risque de s'effondrer sous tes pas, et dans la cave de l'aile gauche où se trouve mon laboratoire. » La voix de l'oncle Vincent, généralement si douce, lui avait paru tellement sévère lors de cette dernière interdiction que jamais il n'avait, malgré sa curiosité, osé enfreindre la consigne. Tous ces souvenirs se pressaient dans sa tête tandis que, d'un pas mal assuré, il s'aventurait vers le sous-sol interdit. « Si au moins j'avais emporté une lampe de poche, » se dit-il lorsque, l'étroit passage obliquant brusquement vers la droite, il se retrouva dans l'obscurité la plus complète. Il allait rebrousser chemin quand soudain il aperçut, juste devant lui, une faible lueur qui lui fit, d'instinct, presser le pas. « Le laboratoire... », murmura-t-il enfin devant une porte massive, légèrement entrouverte, qui laissait filtrer un rayon de lumière crue. Durant quelques secondes, il resta immobile sans oser pousser, ni même toucher le lourd panneau de chêne. Il se décida enfin. Pendant un bref instant, ébloui par la vive clarté de nombreux tubes électriques, il ne put rien discerner. Mais, quand ses yeux se furent habitués, il fut incapable de contenir une brève exclamation à la vue du spectacle qui s'offrait à lui.

§ 2. — **Paragraphe à rédiger par l'élève.**

§ 3. — « Mais non, Monsieur Le Croisic, je vous assure qu'il n'était pas là. J'ai exploré le moindre recoin ; il n'y avait personne. Ce que j'ai fait ? Eh bien ! j'ai d'abord commencé par ouvrir la petite fenêtre du fond. Vous comprenez... l'odeur était insupportable. Et c'est en traversant le laboratoire que j'ai trouvé le bout de papier que

(*) Ce texte a été rédigé par M. G. LEROY, assistant au Séminaire de méthodologie de l'Institut de Psychologie et des Sciences de l'Éducation.

je viens de vous lire. Si c'est bien l'écriture d'oncle Vincent ? Ça, j'en suis sûr ! Vous pensez bien que je la connais, son écriture : depuis que je sais lire, il m'écrit chaque mois... Mais... ! Attendez... C'est vraiment curieux ! Voici seulement que je m'en rends compte : oncle Vincent a fait une faute d'orthographe ! Il a écrit : ... ils sont entré » sans « s » à entré... ! Vous savez comme il est maniaque en ce qui concerne l'ordre, la propreté et surtout... l'orthographe ! Lui, faire une faute pareille ? c'est impensable ! Il n'était certainement pas dans son état normal quand il a écrit cela. Que dites-vous ? Très bien. Donc, vous serez ici dans une vingtaine de minutes. Non, non ! je n'ouvrirai à personne avant votre arrivée... A tout de suite, Monsieur Le Croisic. »

Après avoir raccroché le cornet du téléphone, Luc se rendit à la cuisine pour se laver les mains à grande eau et tenter de faire disparaître le produit malodorant qui lui picotait la peau. « Plus de peur que de mal... » pensa-t-il en constatant que son doigt ne saignait plus et que la coupure était sans gravité. Il but un grand verre de jus de fruit glacé pour se remettre les idées en place et regagna le grand salon toujours aussi désert et silencieux. Il étala le curieux message devant lui, sur la table, le relut attentivement pour la troisième fois, se pencha pour l'observer de plus près, puis, le menton dans le creux de la main, il se mit à réfléchir intensément. Les yeux fixés sur la petite porte sculptée qu'il avait trouvée inexplicablement refermée au retour de son expédition, attentif au moindre bruit, il attendit...

Le travail est pertinent au point de vue des informations appropriées au contexte, c'est-à-dire

B. — Définition des

Deux questions permettront de

- L'élève n'a-t-il pas introduit des informations non pertinentes ? Le travail ne doit pas nécessairement être qu'il ne les contredise pas.
- Les informations utilisées sont-elles pertinentes ? Ces deux aspects de l'évaluation

1^{er} aspect : Introduction d'éléments

Les éléments non pertinents sont des informations apportées par le travail de l'élève. Le travail de l'élève a permis d'établir la liste des informations pertinentes de ce premier aspect de la pertinence.

1. Le début du paragraphe 2 (pousser la porte, entrer, ... sont mentionnés dans le paragraphe 1).
2. Aucune allusion n'est faite au fait que l'enfant est seul, à une exclamation, à un seul mot.
3. Dans le paragraphe 2, le contenu du message est pertinent.
4. L'aspect et le contenu du message deviennent une lettre, longue et précise.
5. L'enfant évoque des blessures.

2^e aspect : Intégration des informations

Le seul repérage des éléments pertinents permet de donner la valeur du travail de l'élève. On doit le faire dans un contexte plus large les

- a) On cherchera les indices de :
 - l'évocation de la personne
 - les allusions à des personnes
 - l'expression des sentiments
 (Cette liste n'est pas exhaustive)
- b) On considérera que l'élève a introduit strictement des faits qui sont pertinents (fin du premier paragraphe) (paragraphe).

PRÉSENTATION DES ÉCHELLES

I. — PERTINENCE SÉMANTIQUE

A. — Définition du critère

Le travail est pertinent au point de vue sémantique si l'élève développe des idées appropriées au contexte, c'est-à-dire au résumé et aux paragraphes 1 et 3.

B. — Définition des aspects qui interviennent dans le critère

Deux questions permettront d'évaluer la pertinence sémantique :

— L'élève n'a-t-il pas introduit d'éléments contredisant les informations fournies ?
Le travail ne doit pas nécessairement comporter toutes ces informations; l'important est qu'il ne les contredise pas.

— Les informations utilisées sont-elles intégrées dans un contexte plus large ?

Ces deux aspects de l'évaluation sont décrits ci-après.

1^{er} aspect : Introduction d'éléments non pertinents.

Les éléments non pertinents sont des éléments qui entrent en contradiction avec les informations apportées par le texte. La lecture d'un nombre important de copies d'élèves a permis d'établir la liste de ces *éléments non pertinents*. Elle servira à évaluer ce premier aspect de la pertinence sémantique.

1. Le début du paragraphe 2 ne situe pas les faits dans le laboratoire (Exemple : pousser la porte, entrer,... sont des actions non pertinentes; elles sont déjà exprimées dans le paragraphe 1).
2. Aucune allusion n'est faite au « spectacle insolite ». Cette allusion peut se limiter à une exclamation, à un seul mot.
3. Dans le paragraphe 2, le commissaire est présent ou attendu.
4. L'aspect et le contenu du message sont modifiés (par exemple, le bout de papier devient une lettre, longue et soignée).
5. L'enfant évoque des blessures graves, des sinistres, des catastrophes,...

2^e aspect : Intégration des informations.

Le seul repérage des éléments non pertinents ne suffit pas à rendre compte de la valeur du travail de l'élève. On appréciera aussi la manière dont l'élève intègre dans un contexte plus large les faits qu'il relate.

a) On cherchera les indices de cette intégration dans :

- l'évocation de la personnalité de l'oncle;
 - les allusions à des personnages secondaires;
 - l'expression des sentiments de Luc,...
- (Cette liste n'est pas exhaustive.)

b) On considérera que l'élève n'a pas intégré les informations s'il se contente de relater strictement des faits qui se situent après l'entrée de Luc dans le laboratoire (fin du premier paragraphe) et avant sa réponse au commissaire (début du troisième paragraphe).

- c) Certains travaux développent uniquement des idées qui, sans être nécessairement en contradiction avec les paragraphes 1 et 3, n'ont pas de rapport avec la situation.

C. — Détermination des niveaux d'exigence

- 1) Pour l'aspect « Introduction d'éléments non pertinents », on retiendra quatre niveaux :
1. Aucun élément non pertinent.
 2. Un ou deux éléments non pertinents.
 3. Trois ou quatre éléments non pertinents.
 4. Plus de quatre éléments non pertinents.
- 2) Pour l'aspect « Intégration », on retiendra trois niveaux :
1. Intégration.
 2. Pas d'intégration.
 3. Rien que des idées sans rapport avec le contexte.

D. — Démarche de l'évaluation

1. Faites porter votre appréciation sur le premier aspect « Introduction d'éléments non pertinents » ⁽¹⁾.

- Aucun élément non pertinent
- 1 ou 2 éléments non pertinents
- 3 ou 4 éléments non pertinents
- Plus de 4 éléments non pertinents

2. Faites porter votre appréciation sur le second aspect « Intégration » ⁽¹⁾.

- L'élève :
- Intègre les informations reçues dans un contexte plus large
 - N'intègre pas les informations
 - Développe uniquement des idées sans rapport avec la situation

3. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux deux caractéristiques que vous venez de repérer dans le travail de l'élève.

<i>Introduction d'éléments non pertinents</i>	<i>Intégration</i>	
— Pas d'éléments non pertinents	— intégration	TB
	— pas d'intégration	B
	— uniquement des idées sans rapport avec la situation	I
— 1 ou 2 éléments non pertinents	— intégration	B
	— pas d'intégration	S
	— uniquement des idées sans rapport avec la situation	I
— 3 ou 4 éléments non pertinents	— qu'il y ait ou non intégration	F
	— uniquement des idées sans rapport avec la situation	I
— Plus de 4 éléments non pertinents		I

Note pour la pertinence sémantique :

⁽¹⁾ Indiquez, par une croix dans la case prévue à cet effet, la réponse que vous choisissez.

	Introduction d'éléments non pertinents
TB	Pas d'éléments non pertinents
B	Pas d'éléments non pertinents
	1 ou 2 éléments non pertinents
S	1 ou 2 éléments non pertinents
F	3 ou 4 éléments non pertinents
I	Plus de 4 éléments non pertinents
	Rien que des idées sans rapport avec la situation

E. — Présentation de l'échelle : « Pertinence sémantique »

	Introduction d'éléments non pertinents	Intégration des informations reçues
TB	Pas d'éléments non pertinents	Intégration des faits du paragraphe 2 dans un contexte plus large
B	Pas d'éléments non pertinents	Pas d'intégration
	1 ou 2 éléments non pertinents	Intégration des faits du paragraphe 2 dans un contexte plus large
S	1 ou 2 éléments non pertinents	Pas d'intégration
F	3 ou 4 éléments non pertinents	Qu'il y ait ou non intégration
I	Plus de 4 éléments non pertinent	Qu'il y ait ou non intégration
	Rien que des idées sans rapport avec la situation, quel que soit le nombre d'éléments non pertinents	

ées qui, sans être nécessairement
t pas de rapport avec la situation.

x d'exigence

ertinents», on retiendra quatre

niveaux :

cte.

uation

aspect « Introduction d'éléments

-
-
-
-

ct « Intégration » (?).

ontexte plus large

port avec la situation

e niveau qui correspond aux deux
travail de l'élève.

on

..... TB
 B
 ns rapport avec la situation I
 B
 S
 ns rapport avec la situation I
 ation F
 ns rapport avec la situation I
 I

la pertinence sémantique :

à cet effet, la réponse que vous

II. — PERTINENCE SYNTAXIQUE

A. — Définition du critère

Respect des contraintes syntaxiques imposées par les paragraphes 1 et 3.

B. — Définition des aspects

Le travail de l'élève est pertinent au point de vue syntaxique s'il concorde avec les temps et les personnes grammaticales utilisés par l'auteur.

1^{er} aspect : *Concordance des temps.*

L'auteur utilise le passé simple, l'imparfait et le plus-que-parfait, ainsi que le présent dans le discours direct.

REMARQUES :

- 1) Les fautes dans la conjugaison des temps ne seront pas sanctionnées (Exemple : Alors, Luc sort...).
- 2) Les erreurs dans la concordance des temps et des modes, et dans le choix des temps ne sont pas pénalisées.

2^e aspect : *La personne grammaticale.*

L'élève est tenu d'utiliser la troisième personne du singulier (première personne du singulier dans le discours direct).

C. — Détermination des niveaux d'exigence

a) Pour l'aspect « Respect des temps », on retiendra trois niveaux :

1. Aucune erreur;
2. Une ou deux erreurs;
3. Plus de deux erreurs.

b) Pour l'aspect « Respect de la personne grammaticale », on retiendra aussi trois niveaux :

1. Aucune erreur;
2. Une ou deux erreurs;
3. Plus de deux erreurs.

D. — Démarche de l'évaluation

1. Faites porter votre appréciation sur le premier aspect : « Concordance des temps ».

- Aucune erreur
- Une ou deux erreurs
- Plus de deux erreurs

2. Faites porter votre appréciation sur le second aspect : « Concordance des personnes grammaticales ».

- Aucune erreur
- Une ou deux erreurs
- Plus de deux erreurs

3. En vous reportant au schéma ci-dessous, indiquez les caractéristiques que vous venez de constater.

Concordance des temps

- Aucune erreur — A
- 1 ou 2 erreurs — B
- Plus de 2 erreurs — C

E. — Présentation

	Concordance des temps
TB	Aucune erreur
B	Aucune erreur
	Une ou deux erreurs
S	Aucune erreur
	Une ou deux erreurs
	Plus de deux erreurs
F	Une ou deux erreurs
	Plus de deux erreurs
I	Plus de deux erreurs

AXIQUE

tière

ar les paragraphes 1 et 3.

aspects

ue syntaxique s'il concorde avec
l'auteur.

plus-que-parfait, ainsi que le pré-

ront pas sanctionnées (Exemple :

des modes, et dans le choix des

du singulier (première personne

x d'exigence

t trois niveaux :

aticale», on retiendra aussi trois

uation

pect : « *Concordance des temps* ».

2. *Faites porter votre appréciation sur le second aspect « Respect de la personne grammaticale ».*

- Aucune erreur
- Une ou deux erreurs
- Plus de deux erreurs

3. *En vous reportant au schéma suivant, choisissez le niveau qui correspond aux deux caractéristiques que vous venez de repérer dans le travail de l'élève.*

<i>Concordance des temps</i>	<i>Respect de la personne grammaticale</i>	
— Aucune erreur	— Aucune erreur	TB
	— 1 ou 2 erreurs	B
	— Plus de 2 erreurs	S
— 1 ou 2 erreurs	— Aucune erreur	B
	— 1 ou 2 erreurs	S
	— Plus de 2 erreurs	F
— Plus de 2 erreurs	— Aucune erreur	S
	— 1 ou 2 erreurs	F
	— Plus de 2 erreurs	I

Note pour la pertinence syntaxique :

E. — *Présentation de l'échelle : « Pertinence syntaxique »*

	Concordance des temps	Respect de la personne grammaticale
TB	Aucune erreur	Aucune erreur
B	Aucune erreur	Une ou deux erreurs
	Une ou deux erreurs	Aucune erreur
S	Aucune erreur	Plus de deux erreurs
	Une ou deux erreurs	Une ou deux erreurs
	Plus de deux erreurs	Aucune erreur
F	Une ou deux erreurs	Plus de deux erreurs
	Plus de deux erreurs	Une ou deux erreurs
I	Plus de deux erreurs	Plus de deux erreurs

III. — VOCABULAIRE

A. — Définition du critère

On évalue l'ensemble des mots utilisés par l'élève dans son texte. On ne tient pas compte de l'orthographe.

B. — Définition des aspects qui interviennent dans le critère

On appréciera le vocabulaire selon un triple point de vue : sa correction, sa précision, son originalité.

1^{er} aspect : *Correction*.

Il est difficile de donner une définition précise de la correction du vocabulaire. On tentera de l'évaluer en repérant les principales incorrections, notamment :

- Les pléonasmes (monter en haut - prévoir d'avance,...);
- Les paronymes (incident/accident - sensé/censé,...);
- Les barbarismes ou néologismes (entièreté - solutionner,...);
- Les belgicismes (chique, s'en aller promener, j'ai bon,...);
- Les solécismes (de suite, parler que,...);
- Extension abusive de sens, glissement de sens (réaliser pour comprendre, savoir pour pouvoir,...);
- Utilisation d'expressions familières inappropriées au contexte;
- Sens inexact, impropriétés de vocabulaire (frappé de « suspect »,...).

2^e aspect : *Précision*.

On sanctionne l'utilisation de :

- Mots inappropriés (Exemple : Il ouvrit la fenêtre pour faire évaporer cette peste);
- Mots passe-partout (Exemples : tube au lieu d'éprouvette; mettre, il y a, faire,...).

On sanctionne aussi la répétition abusive d'un même mot.

3^e aspect : *Originalité*.

Par les deux premiers aspects, on relève les erreurs du travail sans en apprécier les qualités.

Pour nuancer l'évaluation, on prendra en considération l'originalité des mots choisis. On départagera ainsi des travaux comportant un même nombre d'incorrections et d'imprécisions.

C. — Détermination des niveaux d'exigence

a) Pour l'aspect « Correction », on retiendra trois niveaux :

1. Pas d'incorrection;
2. Une ou deux incorrections;
3. Plus de deux incorrections.

b) Pour l'aspect « Précision », on

1. Trois imprécisions maximum;
2. Plus de trois imprécisions.

c) Pour l'aspect « Originalité », o

1. Oui;
2. Non.

D. —

1. Faites porter votre appréciation

- Pas d'incorrection
- Une ou deux incorrections
- Plus de deux incorrections

2. Faites porter votre appréciation

- Trois imprécisions maximum
- Plus de trois imprécisions

3. Faites porter votre appréciation

- Oui
- Non

4. En vous reportant au schéma des caractéristiques que vous venez de

Correction

- Pas d'incorrection { — Trois
- Plus de deux incorrections { — Plus

- Une ou deux incorrections { — Trois
- Plus de deux incorrections { — Plus

- Plus de deux incorrections { — Trois
- Plus de deux incorrections { — Plus

b) Pour l'aspect « Précision », on retiendra deux niveaux :

1. Trois imprécisions maximum;
2. Plus de trois imprécisions.

c) Pour l'aspect « Originalité », on retiendra deux niveaux :

1. Oui;
2. Non.

D. — Démarches de l'évaluation

1. Faites porter votre appréciation sur le premier aspect : « Correction ».

- Pas d'incorrection
- Une ou deux incorrections
- Plus de deux incorrections

2. Faites porter votre appréciation sur le deuxième aspect : « Précision ».

- Trois imprécisions maximum
- Plus de trois imprécisions

3. Faites porter votre appréciation sur le troisième aspect : « Originalité ».

- Oui
- Non

4. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux caractéristiques que vous venez de repérer dans le travail de l'élève.

<i>Correction</i>	<i>Précision</i>	<i>Originalité</i>	
— Pas d'incorrection	<ul style="list-style-type: none"> — Trois imprécisions maximum — Plus de trois imprécisions 	<ul style="list-style-type: none"> — Oui — Non 	TB
			<ul style="list-style-type: none"> — Oui — Non
— Une ou deux incorrections	<ul style="list-style-type: none"> — Trois imprécisions maximum — Plus de trois imprécisions 	<ul style="list-style-type: none"> — Oui — Non 	B
			<ul style="list-style-type: none"> — Oui — Non
— Plus de deux incorrections	<ul style="list-style-type: none"> — Trois imprécisions maximum — Plus de trois imprécisions 	<ul style="list-style-type: none"> — Oui — Non 	S
			<ul style="list-style-type: none"> — Oui — Non

Note pour le vocabulaire :

E. — Présentation de l'échelle « Vocabulaire »

	Correction	Précision	L'élève fait-il preuve d'originalité dans le vocabulaire utilisé?
TB	Pas d'incorrection	3 imprécisions maximum	Oui
B	Pas d'incorrection	3 imprécisions maximum	Non
	Pas d'incorrection	Plus de 3 imprécisions	Oui
	1 ou 2 incorrections	3 imprécisions maximum	Oui
S	Pas d'incorrection	Plus de 3 imprécisions	Non
	1 ou 2 incorrections	3 imprécisions maximum	Non
	1 ou 2 incorrections	Plus de 3 imprécisions	Oui
	Plus de 2 incorrections	3 imprécisions maximum	Oui
F	1 ou 2 incorrections	Plus de 3 imprécisions	Non
	Plus de 2 incorrections	3 imprécisions maximum	Non
	Plus de 2 incorrections	Plus de 3 imprécisions	Oui
I	Plus de 2 incorrections	Plus de 3 imprécisions	Non

Par opposition à la « Pertinence » du graphique rédigé par l'élève, indices graphiques 1 et 3.

B. — Définition de

L'évaluation de la syntaxe s'appuie sur la variété des constructions utilisées.

1^{er} aspect : *Correction*.

On apprécie surtout, à ce stade, la prise comme unité. Beaucoup de constructions attribuables à un mauvais commandement de l'échelle « Organisation des

Il est difficile de définir les erreurs d'une liste exhaustive des incorrections les plus fréquentes :

- Equivoque dans l'utilisation des mots et dans la construction des phrases.
- Rejet en tête de phrase d'un complément.
- Phrases incomplètes (absence de verbe).
- Erreur dans la concordance des temps.
- Raté d'automatisme de la langue (livre que je parle).

2^e aspect : *Variété*.

- Différents registres de variété sont à observer :
- Variété dans les structures syntaxiques : rejets - gallicismes - ellipses.
 - Variété dans les formes utilisées : emploi du style direct.
 - Variété dans la présentation : à l'intérieur d'une même phrase (usage de tirets, parenthèses, ...), entre les phrases : pas de répétitions.
 - Variété dans les procédés d'articulation : complément déterminatif, répétition.

On considérera que l'élève a fait appel à différentes possibilités

Précision	L'élève fait-il preuve d'originalité dans le vocabulaire utilisé?
Précisions maximum	Oui
Précisions maximum	Non
Précisions	Oui
Précisions maximum	Oui
Précisions	Non
Précisions maximum	Non
Précisions	Oui
Précisions maximum	Oui
Précisions	Non
Précisions maximum	Non
Précisions	Oui
Précisions	Non

A. — Définition du critère

Par opposition à la « Pertinence syntaxique », on apprécie ici la syntaxe du paragraphe rédigé par l'élève, indépendamment des contraintes imposées par les paragraphes 1 et 3.

B. — Définition des aspects qui interviennent dans le critère

L'évaluation de la syntaxe se fera à un double point de vue : la correction et la variété des constructions utilisées.

1^{er} aspect : *Correction*.

On apprécie surtout, à ce niveau, la correction syntaxique *interne* de la phrase prise comme unité. Beaucoup d'erreurs dans la « syntaxe entre phrases » sont, en effet, attribuables à un mauvais contrôle sémantique (suite dans les idées) et feront l'objet de l'échelle « Organisation des idées ».

Il est difficile de définir les critères de correction de la phrase française et de dresser une liste exhaustive des incorrections. A titre d'exemple, on retiendra les maladroites les plus fréquentes :

- Equivoque dans l'utilisation des pronoms personnels et des adjectifs possessifs, et dans la construction des relatives;
- Rejet en tête de phrase d'adjectifs ou de participes n'adhérant pas au sujet;
- Phrases incomplètes (absence de terme indispensable à la compréhension);
- Erreur dans la concordance de temps et de modes, et dans le choix du temps;
- Raté d'automatisme de langage : mauvaise utilisation des pronoms (exemple : le livre que je parle).

2^e aspect : *Variété*.

Différents registres de variété peuvent être observés :

- Variété dans les structures : différences dans les attaques de phrases (inversions-rejets - gallicismes - ellipses,...).
- Variété dans les formes utilisées : affirmative, négative, exclamative, interrogative, emploi du style direct.
- Variété dans la présentation des relations possibles :
 - . à l'intérieur d'une même phrase : juxtaposition, coordination, inclusion (emploi de tirets, parenthèses,...), subordination;
 - . entre les phrases : pas de recours abusif aux chevilles (et puis, et alors, et après,...).
- Variété dans les procédés d'expansion du syntagme nominal : épithète, apposition, complément déterminatif, relative.

On considérera que l'élève varie ses phrases dans un registre déterminé si l'élève fait appel à différentes possibilités prévues dans la description de ce registre.

C. — Détermination des niveaux d'exigence

a) Pour l'aspect « Correction des structures », on retiendra trois niveaux :

1. Pas d'incorrection
2. Une ou deux incorrections
3. Plus de deux incorrections.

b) Pour l'aspect « Variété de structures », on retiendra quatre niveaux :

1. Trois registres de variété au moins
2. Deux registres de variété
3. Un seul registre
4. Pas de variété.

D. — Démarche de l'évaluation du critère : « Structures grammaticales »

1. Faites porter votre appréciation sur le premier aspect : « Correction ».

- Pas d'incorrection
- Une ou deux incorrections
- Plus de deux incorrections

2. Faites porter votre appréciation sur le deuxième aspect : « Variété ».

- Trois registres de variété au moins
- Deux registres de variété
- Un seul registre
- Pas de variété

3. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux deux caractéristiques que vous venez de repérer dans le travail de l'élève.

- | | | | |
|------------------------------|---|--|----|
| — Aucune incorrection | { | — Trois registres de variété au moins .. | TB |
| | | — Deux registres de variété | B |
| | | — Un seul registre | S |
| | | — Pas de variété | F |
| — Une ou deux incorrections | { | — Trois registres de variété au moins .. | B |
| | | — Deux registres de variété | S |
| | | — Un seul registre | F |
| | | — Pas de variété | I |
| — Plus de deux incorrections | { | — Trois registres de variété au moins .. | S |
| | | — Deux registres de variété | F |
| | | — Un seul registre | I |
| | | — Pas de variété | I |

Note pour les structures grammaticales :

E. — Présentation

	Correct
TB	Aucune incorrection
B	Aucune incorrection
	Une ou deux incorrections
S	Aucune incorrection
	Une ou deux incorrections
	Plus de deux incorrections
F	Aucune incorrection
	Une ou deux incorrections
	Plus de deux incorrections
I	Une ou deux incorrections
	Plus de deux incorrections
	Plus de deux incorrections

aux d'exigence

retiendra trois niveaux :

dra quatre niveaux :

« Structures grammaticales »

pect : « Correction ».

spect : « Variété ».

le niveau qui correspond aux deux
le travail de l'élève.

tres de variété au moins .. TB
tres de variété B
gistre S
été F

tres de variété au moins .. B
tres de variété S
gistre F
été I

tres de variété au moins .. S
tres de variété F
gistre I
été I

s structures grammaticales :

E. — Présentation de l'échelle : « Structures grammaticales »

	Correction	Variété
TB	Aucune incorrection	Trois registres de variété au moins
B	Aucune incorrection	Deux registres de variété
	Une ou deux incorrections	Trois registres de variété au moins
S	Aucune incorrection	Un seul registre de variété
	Une ou deux incorrections	Deux registres de variété
	Plus de deux incorrections	Trois registres de variété au moins
F	Aucune incorrection	Pas de variété
	Une ou deux incorrections	Un seul registre de variété
	Plus de deux incorrections	Deux registres de variété
I	Une ou deux incorrections	Pas de variété
	Plus de deux incorrections	Un seul registre de variété
	Plus de deux incorrections	Pas de variété

V. — ORGANISATION DES IDÉES DANS LE PARAGRAPHE 2

A. — Définition du critère

Par opposition avec la dimension « Pertinence sémantique », on évalue la structuration du travail, *indépendamment du respect des contraintes imposées par les paragraphes 1 et 3.*

B. — Description des aspects qui interviennent dans le critère

Cette évaluation portera sur trois aspects : la cohérence sémantique, la liaison des idées et la progression du récit.

1. La cohérence sémantique.

Les idées exprimées par l'élève ne doivent pas présenter de contradiction entre elles.

2. La liaison des idées.

On s'attache à évaluer la manière dont l'élève lie les idées les unes aux autres, soit par l'utilisation de liens (conjonctions de coordination, locutions adverbiales, etc.), soit par la ponctuation.

L'utilisation de liens inadéquats sera sanctionnée. Exemples :

— Absence d'un lien nécessaire : « Il le ramassa et se coupa quelque chose était écrit. »

— Choix inadéquat de ce lien : « Il alla dans une autre chambre qu'il ne connaissait pas et c'était la cuisine. » — « Il sentit des odeurs de produit chimique car il y avait un désordre. »

De même, des erreurs de ponctuation flagrantes seront pénalisées dans la mesure où elles entraînent une liaison maladroite des idées.

3. La progression du récit.

On évalue la manière dont l'élève donne un rythme soutenu à son texte.

La présence d'éléments redondants qui ralentissent la progression du récit est pénalisée. Exemple : « En ouvrant la grande porte, il se blessa légèrement au doigt. Lorsqu'il eut ouvert la grande porte, Luc monta l'escalier. »

C. — Détermination des niveaux d'exigence

a) Pour l'aspect « Cohérence sémantique », on retiendra trois niveaux :

1. Récit cohérent
2. Contradictions passagères
3. Récit incohérent.

- b) Pour l'aspect « Liaison des idées » :
1. Liaison généralement adroite
 2. Liaison maladroite.

- c) Pour l'aspect « Progression du récit » :
1. Récit non redondant
 2. Progression ralentie.

D. — Démarche de l'évaluation

1. Faites porter votre appréciation sur :
- Récit cohérent
 - Contradictions passagères
 - Récit incohérent

2. Faites porter votre appréciation sur :
- Liaison généralement adroite
 - Liaison maladroite

3. Faites porter votre appréciation sur :
- Récit non redondant
 - Progression ralentie

4. En vous reportant au schéma ci-dessous, indiquez les caractéristiques que vous observez :

Cohérence sémantique	Liaison des idées
— Récit cohérent	— Liaison généralement adroite
	— Liaison maladroite
	— Récit incohérent
— Contradictions passagères	— Liaison généralement adroite
	— Liaison maladroite
— Récit incohérent	— Liaison généralement adroite
	— Liaison maladroite

critère

« Cohérence sémantique », on évalue la structure sémantique imposée par les paramètres

« Liaison des idées », on évalue la liaison des idées

« Progression du récit », on évalue la progression du récit

présenter de contradiction entre les idées les unes aux autres, soit par des locutions adverbiales, etc.),

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

Exemples :
« Il se coupa quelque chose était dans la chambre qu'il ne connaissait pas de produit chimique car il y avait des produits dangereux qui seraient pénalisés dans la mesure où ils sont soutenus à son texte. La progression du récit est pénalisée légèrement au doigt. Lors de la lecture. »

b) Pour l'aspect « Liaison des idées », on retiendra deux niveaux :

1. Liaison généralement adroite
2. Liaison maladroite.

c) Pour l'aspect « Progression du récit », on retiendra deux niveaux :

1. Récit non redondant
2. Progression ralentie.

D. — Démarche de l'évolution du critère : « Organisation des idées dans le paragraphe 2 »

1. Faites porter votre appréciation sur le premier aspect : « Cohérence sémantique ».

- Récit cohérent
- Contradictions passagères
- Récit incohérent

2. Faites porter votre appréciation sur le second aspect : « Liaison des idées ».

- Liaison généralement adroite
- Liaison maladroite

3. Faites porter votre appréciation sur le troisième aspect : « Progression du récit ».

- Récit non redondant
- Progression ralentie

4. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux trois caractéristiques que vous venez de repérer dans le travail de l'élève.

<i>Cohérence sémantique</i>	<i>Liaison des idées</i>	<i>Progression du récit</i>
— Récit cohérent	<ul style="list-style-type: none"> — Liaison généralement adroite — Liaison maladroite 	<ul style="list-style-type: none"> — Récit non redondant . TB — Progression ralentie .. B — Récit non redondant . B — Progression ralentie .. S
— Contradictions passagères	<ul style="list-style-type: none"> — Liaison généralement adroite — Liaison maladroite 	<ul style="list-style-type: none"> — Récit non redondant . B — Progression ralentie .. S — Récit non redondant . S — Progression ralentie .. F
— Récit incohérent	<ul style="list-style-type: none"> — Liaison généralement adroite — Liaison maladroite 	<ul style="list-style-type: none"> — Récit non redondant . F — Progression ralentie .. I — Récit non redondant . I — Progression ralentie .. I

Note pour l'organisation des idées dans le paragraphe 2 :

	Cohérence sémantique	Liaison des idées	Progression du récit
TB	Récit cohérent	Liaison généralement adroite	Récit non redondant
B	Récit cohérent	Liaison généralement adroite	Progression ralentie
	Récit cohérent	Liaison maladroite	Récit non redondant
	Contradictions passagères	Liaison généralement adroite	Récit non redondant
S	Récit cohérent	Liaison maladroite	Progression ralentie
	Contradictions passagères	Liaison généralement adroite	Progression ralentie
	Contradictions passagères	Liaison maladroite	Récit non redondant
F	Contradictions passagères	Liaison maladroite	Progression ralentie
	Récit incohérent	Liaison généralement adroite	Récit non redondant
I	Récit incohérent	Liaison généralement adroite	Progression ralentie
	Récit incohérent	Liaison maladroite	Récit non redondant
	Récit incohérent	Liaison maladroite	Progression ralentie

Vous venez d'attribuer une note

Reportez ces notes sur les échell

1. Pertinence sémantique

2. Pertinence syntaxique

3. Vocabulaire

4. Structures grammaticales

5. Organisation des idées

Note de synthèse

Idées	Progression du récit
ment	Récit non redondant
ment	Progression ralentie
ite	Récit non redondant
ment	Récit non redondant
ite	Progression ralentie
ment	Progression ralentie
ite	Récit non redondant
ite	Progression ralentie
ment	Récit non redondant
ment	Progression ralentie
ite	Récit non redondant
ite	Progression ralentie

Vous venez d'attribuer une note pour chacun des critères.

Reportez ces notes sur les échelles ci-après :

1. Pertinence sémantique

TB	B	S	F	I

2. Pertinence syntaxique

TB	B	S	F	I

3. Vocabulaire

TB	B	S	F	I

4. Structures grammaticales

TB	B	S	F	I

5. Organisation des idées

TB	B	S	F	I

Note de synthèse

TB	B	S	F	I

Christine J. § II. Une aventure de Luc.

Des tas d'objets éparpillés, tubes en verre, qui ~~contenaient~~ ^{contenaient} de toutes sortes de produits, une paire de ciseaux ... etc... tout cela sur une grande table de travail. "oh, là là quel désordre!" s'exclama Luc. Ah! quel odeur il ne pouvait plus respirer. Il traversa le laboratoire se courrant pour vite ouvrir la petite fenêtre. Cela fait du bien. Mais en courant il avait entreprit un bout de papier - Luc se pencha pour le ramasser et il le lut "Ils sont entrés..." Après avoir lut et relut le morceau de papier, il chercha après son oncle vraiment il avait beau chercher pas d'oncle. Luc voulait le trouver. Luc faire. Il n'était pas très rassuré. Il serait ~~par~~ son cœur battre à toute allure. Pas d'oncle, pas de Georges, pas d'Anna. Luc était au bord des larmes. Doudoir de bruits légers se firent entendre. Luc se décida à monter les bruits étaient de plus en fort. Il redoublait de peur. Il se dirigea vers la cuisine c'était de là que venait les bruits - le robinet! oui le robinet était ouvert. Les bruits c'était donc cela - Luc se dit: que faire mais voyons! je vais tout de suite téléphoner à M^r Le Croisic.

Ding Ding "Allo bonjour monsieur je vient ~~est~~ ~~est~~ (parler) d'arrivé chez mon oncle je ne le vois pas, mais voyons il doit être là ... mais non..."

LANG

Enseigne

Éc

RÉSUMÉ E

Ce dossier est dû à la collabora

Mmes Micheline DE JAEGER, Ath
Germaine LAMBION, Lycée
Françoise MELIGNON, Ath

MM. Philippe CARON, Universit
Yvon DELEPIERRE, Athén
Raoul MELIGNON, Athén
Jean TRIGAUX, Athénée R

Chercheurs : Jacqueline PAQUAY

qui ^{contenait} ~~contenait~~ de
ciseau ... etc... tout
oh, la la quel
l'odeur il ne pouvait
re en courant pour
ait du lien. Mais
de papier - Luc se
ait " Ils sont entré .."
de papier, il
vait beau chercher pas
e faire. Il n'était
cœur battre à
eorge, pas d'Anna
lair des bruits légers
à monter les bruits
t de peur. Il se
la que venait la
ait ouvert. Les bruits
faire mais voyons!
de Croisic.
sieur je vent ~~est~~
de je ne le vois pas,
non...

LANGUE MATERNELLE

Enseignement secondaire rénové : 2^e cycle

ÉCHELLES D'ÉVALUATION

RÉSUMÉ ET CRITIQUE D'UN TEXTE

Ce dossier est dû à la collaboration d'une commission de professeurs composée de :

Mmes Micheline DE JAEGER, Athénée Royal de Gembloux;
Germaine LAMBION, Lycée Royal d'Etterbeek;
Françoise MELIGNON, Athénée Royal de Liège II.

MM. Philippe CARON, Université de Nancy II;
Yvon DELEPIERRE, Athénée Royal de Saint-Hubert;
Raoul MELIGNON, Athénée Royal de Jupille;
Jean TRIGAUX, Athénée Royal de Marche.

Chercheurs : Jacqueline PAQUAY-BECKERS et Robert DE BAL.

L'évaluation fait partie intégrante du processus d'enseignement dont elle est une composante essentielle. Aussi importe-t-il de garantir la *validité* des instruments de mesure utilisés. Un instrument valide est représentatif de la matière dont on souhaite vérifier l'apprentissage : les questions explorent à la fois tous les contenus et tous les comportements applicables à ces contenus. Une telle procédure nécessite des études théoriques qui recensent les éléments fondamentaux de l'acquisition d'une notion. L'évaluation permet alors d'informer chaque élève de son progrès vers un objectif déterminé. En outre, la performance de l'élève peut être comparée à celles d'une population de référence (étalonnage normatif). Nous sommes encore loin de disposer des études et des instruments nécessaires...

Pour remplir son rôle vis-à-vis de l'élève, l'évaluation doit être aussi *objective* et *nuancée* que possible. Le degré de concordance entre différents juges est un indice important de l'objectivité d'une note. Or de nombreuses études docimologiques ont montré, pratiquement pour toutes les disciplines, combien les juges s'accordent peu sur la note à attribuer à une performance donnée. De plus, l'élève ne peut tirer un profit réel de l'évaluation que si les informations qui l'accompagnent sont analytiques, complètes et précises. Pour que le professeur puisse apporter un tel *feedback* à l'élève, la démarche de la note doit, elle aussi, être analytique.

Les difficultés de l'évaluation sont d'autant plus grandes que la performance sollicitée de l'élève nécessite la mise en œuvre de connaissances et d'aptitudes nombreuses. Les échelles descriptives présentées ici ont été construites pour l'évaluation de performances complexes. Des professeurs ont mis leur expérience en commun pour choisir des critères d'appréciation aussi sûrs, aussi valides que possible. Ces critères ont été appliqués à de nombreux travaux.

Dans l'espoir d'assurer l'homogénéité des notes, des consignes de correction détaillées sont proposées (et justifiées).

L'évaluation à l'aide d'échelles descriptives fournit deux types d'information à l'élève :

- Il est informé des points forts et des lacunes de son travail;
- Il dispose de points de repère pour situer sa performance par rapport à ses condisciples.

En raison de son caractère analytique, l'évaluation à l'aide d'échelles descriptives prend beaucoup de temps : elle ne pourrait pas être pratiquée quotidiennement dans une classe. Elle devrait être réservée pour certains moments cruciaux de l'apprentissage, ou, plus simplement, pour des vérifications périodiques. Au lieu de se servir des batteries complètes qui lui sont proposées, le professeur a aussi toujours le loisir de n'utiliser qu'une ou deux des échelles, selon les besoins.

On peut espérer que, petit à petit, se créeront des banques d'instrument d'évaluation auxquelles les enseignants pourront faire appel à leur convenance et à celle de leurs élèves.

Notes pour l'évaluation des travaux.

1. Les critères d'évaluation proposés dans les échelles ci-après ont été définis et testés par la commission de professeurs, au cours de nombreuses séances de travail. Il se peut que ces critères ne soient pas ceux en fonction desquels vous jugez habituellement. Pour essayer l'instrument, il importe cependant que vous vous conformiez strictement aux consignes de correction proposées.

En vue de modifications ultérieures précieuses.

2. Pour la commodité et l'efficacité des travaux de vos élèves selon la suite des pages suivantes.

d'enseignement dont elle est une
tir la *validité* des instruments de
tif de la matière dont on souhaite
la fois tous les contenus et tous
ne telle procédure nécessite des
amentaux de l'acquisition d'une
que élève de son progrès vers un
élève peut être comparée à celles
if). Nous sommes encore loin de

évaluation doit être aussi *objective*
entre différents juges est un indice
reuses études docimologiques ont
ombien les juges s'accordent peu
De plus, l'élève ne peut tirer un
l'accompagnement sont analytiques,
apporter un tel *feedback* à l'élève,
rique.

grandes que la performance solli-
sances et d'aptitudes nombreuses.
construites pour l'évaluation de
leur expérience en commun pour
valides que possible. Ces critères

es, des consignes de correction

rnit deux types d'information à

le son travail;
performance par rapport à ses

on à l'aide d'échelles descriptives
e pratiquée quotidiennement dans
ments cruciaux de l'apprentissage,
ques. Au lieu de se servir des bat-
a aussi toujours le loisir de n'uti-

s banques d'instrument d'évalua-
l à leur convenance et à celle de

es ci-après ont été définis et testés
nombreuses séances de travail. Il
tion desquels vous jugez habituel-
endant que vous vous conformiez
es.

En vue de modifications ultérieures des critères, vos remarques et vos critiques seront précieuses.

2. Pour la commodité et l'efficacité de l'évaluation, il est conseillé d'évaluer tous les travaux de vos élèves selon la première échelle, puis de procéder de même pour les suivantes.

A. — Description de l'exercice choisi

En début d'année scolaire, des élèves de quatrième année de l'enseignement secondaire rénové ont été invités à résumer et à critiquer les idées d'un extrait des *Propos sur l'éducation d'Alain*.

Le texte, tel qu'il figure à la page 3, a été lu individuellement. De manière à ne pas influencer la performance de l'élève, le texte a été présenté sans titre. Le professeur n'a fourni aucune information complémentaire.

Les tâches demandées étaient les suivantes :

1. Résume les idées du texte le plus brièvement possible, en te servant de tes propres mots.
2. Discute les idées émises par l'auteur, en te servant de tes propres arguments.

Les deux tâches portent sur le même texte et ont été présentées simultanément aux élèves. Elles constituent cependant des performances distinctes. C'est pourquoi deux dossiers différents seront présentés ci-après : le premier sera consacré à l'évaluation du résumé, le second à la critique.

Il y a un enseignement monarchique, j'entends un enseignement qui a pour objet de séparer ceux qui sauront et gouverneront de ceux qui ignoreront et obéiront.

Je revois par l'imagination notre professeur de mathématiques; je le revois, écrasant de son ironie un peu lourde un de nos camarades, qui était aussi myope qu'on peut l'être. Cet enfant ne voyait qu'au bout de son nez. Aussi promenait-il son nez d'un bout de la ligne à l'autre, afin de s'en donner une perception exacte; quant à voir le triangle tout entier, d'un seul regard, il n'y pouvait point songer. Je suppose qu'il aurait fallu l'exercer sur de toutes petites figures, pas plus larges que le bout de son nez; ainsi, découvrant le triangle tout entier, il aurait pu y saisir des rapports, et raisonner après cela aussi bien qu'un autre.

Mais il s'agissait bien de cela. On le pressait. Il courait d'un sommet du triangle à l'autre, parlait pour remplir le temps, disait A pour B, droite pour angle, ce qui faisait des discours parfaitement ridicules, et nous avions des rires d'esclaves. Cet enfant fut ainsi condamné publiquement à n'être qu'un sot, parce qu'il était myope.

Cet écrasement des faibles exprime tout un système politique dans lequel nous sommes encore à moitié empêtrés. Il semble que le professeur ait pour tâche de choisir, dans la foule, une élite, et de décourager et rabattre les autres...

Selon mon idée, il faudrait agir tout à fait autrement : instruire le peuple tout entier; se plier à la myopie, à la lourdeur d'esprit, aiguillonner la paresse, éveiller à tout prix ceux qui dorment, et montrer plus de joie pour un petit paysan un peu débarbouillé que pour un élégant mathématicien qui s'élève d'un vol sûr jusqu'aux sommets de l'Ecole Polytechnique (1). D'après cela, tout l'effort des Pouvoirs Publics devrait s'employer à éclairer les masses par dessous et par le dedans, au lieu de faire briller quelques pics superbes, quelques rois nés du peuple, et qui donnent un air de justice à l'inégalité.

ALAIN, *Propos* (Gallimard)

(1) Ecole Polytechnique : Une des écoles supérieures françaises les plus réputées.

B. — Justifi

L'échelle d'évaluation est un in-
plexes pour lesquelles un simple d
Parmi les performances complexes
secondaire supérieur, résumer et c
sage fondamental.

C. —

Chaque dossier comprend :

- La définition du critère;
- La description des aspects qui
- La détermination des niveaux
- La démarche de l'évaluation;
- La présentation de l'échelle.

Les exigences, pour chaque nri
été établies après une analyse app
de quatrième année de l'enseign

me année de l'enseignement second-
er les idées d'un extrait des *Propos*

viduellement. De manière à ne pas
présenté sans titre. Le professeur

ossible, en te servant de tes propres

servant de tes propres arguments.

été présentées simultanément aux
es distinctes. C'est pourquoi deux
emier sera consacré à l'évaluation

enseignement qui a pour objet de
qui ignoreront et obéiront.

thématiques; je le revois, écrasant
était aussi myope qu'on peut l'être.
romenait-il son nez d'un bout de la
xacte; quant à voir le triangle tout
Je suppose qu'il aurait fallu l'exer-
e bout de son nez; ainsi, découvrant
orts, et raisonner après cela aussi

courait d'un sommet du triangle à
B, droite pour angle, ce qui faisait
les rires d'esclaves. Cet enfant fut
ce qu'il était myope.

p politique dans lequel nous sommes
r ait pour tâche de choisir, dans la
res...

nt : instruire le peuple tout entier;
ner la paresse, éveiller à tout prix
tit paysan un peu débarbouillé que
sûr jusqu'aux sommets de l'École
ouvoirs Publics devrait s'employer
u lieu de faire briller quelques pics
nt un air de justice à l'inégalité.

ALAIN, *Propos* (Gallimard)

eurs françaises les plus réputées.

B. — Justification du choix de cet exercice

L'échelle d'évaluation est un instrument destiné à évaluer des performances complexes pour lesquelles un simple décompte de points par faute commise est inadéquat. Parmi les performances complexes que l'on peut envisager au début de l'enseignement secondaire supérieur, résumer et critiquer un texte est, de l'avis général, un apprentissage fondamental.

C. — Organisation des dossiers

Chaque dossier comprend :

- La définition du critère;
- La description des aspects qui interviennent dans l'évaluation;
- La détermination des niveaux d'exigence;
- La démarche de l'évaluation;
- La présentation de l'échelle.

Les exigences, pour chaque niveau de l'échelle, ne sont pas arbitraires : elles ont été établies après une analyse approfondie d'un nombre important de travaux d'élèves de quatrième année de l'enseignement secondaire rénové.

I. — RÉSUMÉ

1. — DÉFINITION DU CRITÈRE.

Résumer, c'est restituer, sous forme synthétique, les idées maîtresses d'un texte.

2. — DESCRIPTION DES ASPECTS.

Cette définition suggère les deux aspects à évaluer :

- La compréhension des idées principales du texte;
- La concision dans leur présentation.

Pour l'attribution de la note relative au résumé, ces deux aspects auront un poids égal.

L'évaluation ne portera pas sur l'orthographe, le vocabulaire, les structures grammaticales, etc. L'exercice proposé n'est pas spécialement destiné à mesurer un apprentissage dans ces domaines.

I. — Compréhension des idées principales du texte

Le texte d'Alain contient *trois idées principales* :

- A. L'enseignement actuel revêt un caractère monarchique, aristocratique.
- B. Il est nécessaire de passer à un enseignement populaire.
- C. Il existe une relation entre l'enseignement et le système politique, et l'on fait des concessions à ce système (« donner un air de justice à l'inégalité »).

La compréhension est évaluée à un double point de vue :

- Est-elle complète ?
- Est-elle correcte ?

a) Compréhension complète des idées principales.

L'idée A comprend les éléments suivants :

1. Il s'agit d'un enseignement Il suffit que l'élève mentionne une seule fois dans sa réponse qu'il s'agit d'enseignement.
2. qui favorise L'élève doit faire apparaître l'idée d'une différence de traitement en faveur de certains élèves au détriment des autres...
3. une élite (ceux qui sauront et gouverneront) par rapport à la masse (ceux qui ignoreront et obéiront). L'élève doit se dégager du cas particulier présenté par Alain (le myope; avoir des préférences pour un élève...) et faire apparaître qu'il s'agit d'un système d'enseignement.

L'idée B comprend les éléments suivants :

1. Il faut passer L'élève doit faire apparaître l'idée d'un changement nécessaire. Cette idée peut éventuellement s'exprimer par l'emploi d'un simple conditionnel.

2. à un enseignement

3. qui donne des chances égales à t

Comme on le remarquera dans 1 et B.

L'idée C :

L'analyse d'un grand nombre de très rare au 2^e cycle de l'enseignement cette idée n'intervient pas dans l'évaluation d'une bonne performance p témoigne d'une incompréhension pénalisée.

Pour chacune des idées,

- L'élève ne fait preuve de compréhension de tous les éléments énoncés ci-dessus.
- Sa compréhension est *partielle*.
- Il y a *omission* si tous les éléments

b) Compréhension correcte des idées

On pénalisera la présence d'idées

Une idée faussée est une idée qui ne correspond pas à ce qu'Alain.

Ex. : « Il faut changer l'enseignement »

Parfois, le professeur peut considérer que l'élève a *remette son évaluation en cause* quand l'élève reprend les mots de l'énoncé dans une formulation personnelle.

EN RÉSUMÉ :

Il s'agit de déterminer si chaque

- Complète et correctement comprise;
- Partielle et correctement comprise;
- Faussée;
- Omise.

La consigne impose à l'élève de faire apparaître les idées principales du texte.

Pour cette raison, on pénalise to

2. à un enseignement L'élève ne doit pas obligatoirement répéter cet élément de la réponse si celui-ci figure déjà dans l'expression de l'idée A.
3. qui donne des chances égales à tous.. Si Alain défend un enseignement, non pas identique pour tous, mais *adapté aux besoins de chacun*, on se contentera d'une précision moindre chez l'élève de l'enseignement secondaire et on considérera comme correcte l'idée d'un enseignement *égalitaire*.

Comme on le remarquera dans l'échelle, on accordera le même poids aux idées A et B.

L'idée C :

L'analyse d'un grand nombre de copies montre que la restitution de cette idée est très rare au 2^e cycle de l'enseignement rénové. C'est pourquoi la compréhension de cette idée n'intervient pas dans l'échelle; éventuellement, le professeur peut cependant valoriser une bonne performance par un commentaire approprié. Si le travail de l'élève témoigne d'une incompréhension totale ou partielle de cette idée, il ne sera pas pénalisé.

Pour chacune des idées,

- L'élève ne fait preuve de compréhension *complète* que si sa réponse comporte tous les éléments énoncés ci-dessus;
- Sa compréhension est *partielle* si un des éléments manque;
- Il y a *omission* si tous les éléments manquent.

b) Compréhension correcte des idées principales.

On pénalisera la présence d'*idées faussées*.

Une idée faussée est une idée du texte à laquelle l'élève attribue un autre sens qu'Alain.

Ex. : « Il faut changer l'enseignement afin de permettre à tous de gouverner. »

Parfois, le professeur peut considérer que l'élève *comprend* une idée du texte, puis *remettre son évaluation en cause* en lisant la suite du travail. Ce cas se présente surtout quand l'élève reprend les mots de l'auteur pour exprimer une idée et introduit ensuite dans une formulation personnelle, une idée faussée ou incompréhensible.

EN RÉSUMÉ :

Il s'agit de déterminer si chaque idée est :

- Complète et correctement comprise;
- Partielle et correctement comprise;
- Faussée;
- Omise.

II. — Concision

La consigne impose à l'élève de résumer « le plus brièvement possible » les idées principales du texte.

Pour cette raison, on pénalise toute répétition d'*idée* (complète, partielle ou faussée),

sous quelque forme que ce soit (mots du texte ou paraphrase), pour autant que l'idée en question ait été exprimée une première fois.

On ne pénalise pas la répétition de *mots* ou de *groupes de mots* nécessaires pour éviter les ambiguïtés.

EXEMPLES :

Elève A : Si mes idées étaient appliquées, l'enseignement, de monarchique, deviendrait démocratique; c'est-à-dire que tout le monde aurait la même chance, aussi bien un petit paysan que le fils d'une famille très aisée: de cette façon, les différences entre les couches sociales auraient tendance à disparaître petit à petit, et le peuple sortirait de l'ignorance où il est encore et s'élèverait lui aussi vers les sommets des grandes écoles.

Jusque « très aisée », le texte de l'élève prouve qu'il a compris les idées A et B. La suite est considérée comme redondante.

Elève B : L'enseignement monarchique qui humilie et écrase la masse devrait être remplacé par un enseignement qui mette tout le monde sur le même pied.

Les verbes *humilier* et *écraser* sont des variations stylistiques qui ne seront pas pénalisées.

3. — DÉTERMINATION DES NIVEAUX D'EXIGENCE.

a) Pour l'aspect « *Compréhension des idées principales du texte* », on retiendra cinq niveaux d'exigence.

Le travail comporte :

1. Deux idées complètes;
2. Une idée complète et une idée partielle;
3. Une idée complète et une idée faussée ou omise;
ou deux idées partielles;
4. Une idée partielle et une idée faussée ou omise;
5. Deux idées faussées ou omises.

b) Pour l'aspect « *Concision* », on ne retiendra que deux niveaux :

1. Le travail est concis;
2. Le travail n'est pas concis; il présente au moins un cas de redondance.

4. — DÉMARCHE POUR L'ÉVALUATION DU RÉSUMÉ.

1. *Faites porter votre appréciation sur le premier aspect : « Compréhension ».*

— L'idée A est-elle :

- . complètement comprise ?
- . partiellement comprise ?
- . faussée ?
- . omise ?

— L'idée B est-elle :

- . complètement comprise ?
- . partiellement comprise ?
- . faussée ?
- . omise ?

Choisissez le niveau qui correspond à votre appréciation.
Le travail comporte-t-il :

- Deux idées complètes ?
- Une idée complète et une idée partielle ?
- Une idée complète et une idée faussée ou omise ?
- Une idée partielle et une idée faussée ou omise ?
- Deux idées faussées ou omises ?

2. *Faites porter votre appréciation sur la concision :*

- Le travail est-il concis ?

3. *En vous reportant au schéma suivant, indiquez les caractéristiques que vous venez de relever :*

1^{er} aspect : *Compréhension des idées principales A et B*

- Deux idées complètes

- Une idée complète et une idée partielle

- Une idée complète et une idée faussée ou omise

- Deux idées partielles

- Une idée partielle et une idée faussée ou omise

- Deux idées faussées ou omises

5. — PRÉSENTATION DE L'ÉCHELLE.

Pour formuler une évaluation, comparez le résumé à la démarche recommandée. Indiquez l'avantage de donner une description

paraphrase), pour autant que l'idée
groupes de mots nécessaires pour éviter

nement, de monarchique, deviendrait
monde aurait la même chance, aussi
mille très aisée: de cette façon, les
raient tendance à disparaître petit à
ce où il est encore et s'élèverait lui

proouve qu'il a compris les idées
redondante.

ie et écrase la masse devrait être
e tout le monde sur le même pied.
variations stylistiques qui ne seront

ipales du texte», on retiendra cinq

mise;

mise;

deux niveaux :

moins un cas de redondance.

aspect : « Compréhension ».

Choisissez le niveau qui correspond aux caractéristiques que vous venez de repérer.
Le travail comporte-t-il :

- Deux idées complètes ?
- Une idée complète et une idée partielle ?
- Une idée complète et une idée faussée ou omise ?
- ou : deux idées partielles ?
- Une idée partielle et une idée faussée ou omise ?
- Deux idées faussées ou omises ?

2. Faites porter votre appréciation sur le deuxième aspect : « Concision ».

- Le travail est-il concis ?
- OUI
- NON

3. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux trois caractéristiques que vous venez de repérer dans le travail de l'élève.

1 ^{er} aspect : Compréhension des idées principales A et B	2 ^e aspect : Concision	Niveaux de l'échelle
— Deux idées complètes	{ — Oui — Non	TB B
— Une idée complète et une idée partielle	{ — Oui — Non	B S
— Une idée complète et une idée faussée ou omise ou	} — Oui — Non	S F
— Deux idées partielles		
— Une idée partielle et une idée faussée ou omise	} — Oui — Non	F I
— Deux idées faussées ou omises		

Note pour le résumé :

5. — PRÉSENTATION DE L'ÉCHELLE.

Pour formuler une évaluation, l'échelle n'apporte aucune information complémentaire par rapport à la démarche décrite précédemment. L'échelle offre toutefois l'avantage de donner une description hiérarchique de la performance.

	1 ^{er} aspect : Compréhension	2 ^e aspect : Concision
TB	Deux idées complètes	Oui
B	Deux idées complètes	Non
	Une idée complète et une idée partielle	Oui
S	Une idée complète et une idée partielle	Non
	Une idée complète et une idée faussée ou omise <i>Ou</i> : Deux idées partielles	Oui
F	Une idée complète et une idée faussée ou omise <i>Ou</i> : Deux idées partielles	Non
	Une idée partielle et une idée faussée ou omise	Oui
I	Une idée partielle et une idée faussée ou omise	Non
	Deux idées faussées ou omises	quelles que soient les caractéristiques de concision

1. — DÉFINITION DU CRITÈRE.

Critiquer, c'est faire l'examen de d'elles.

2. — DESCRIPTION DES ASPECTS.

Cette définition impose les deux :

- Le nombre et la correction des idées
- Le type et la validité des arguments

L'échelle est conçue de façon à être utilisée et à leur validité, qu'au argumentation est, en effet, l'essentiel.

Comme pour le résumé, on n'évalue pas les idées faussées ou omises.

En outre, on n'a pas prévu d'échelle pour les arguments. Il sera jugé en même temps.

I. — Nombre et correction

On n'évalue ici que les idées développées.

1. Seules les critiques portant sur les idées développées et non celles qui sont relatives au résumé.
2. L'élève ne critique pas :
 - si son travail reprend une idée développée, sans le justifier d'aucune façon ;
 - s'il se contente de paraphraser.
3. On ne retiendra dans l'évaluation que les arguments valides. On préconisera au paragraphe.

II. — Type et validité

Cette évaluation se fera en deux étapes :

- a) On donnera tout d'abord un avis sur la validité des arguments.

Il n'appartient pas au professeur de juger la validité des arguments avancés par l'élève pour le simple fait qu'ils sont retenus seront d'ordre logique.

Est tenu pour valide un argument :

- ni erreur de raisonnement (*),
- ni recours à des arguments non pertinents (sans rapport avec le principe à démontrer),
- ni pétitions de principe (reprendre ce qu'il faut démontrer).

(*) Cette liste n'est pas destinée à juger la validité des arguments que contient le travail d'un élève. Elle est destinée au professeur à déterminer si un argument est valide.

on	2 ^e aspect : Concision
	Oui
	Non
partielle	Oui
partielle	Non
faussée	Oui
les	
faussée	Non
les	
faussée	Oui
faussée	Non
	quelles que soient les caractéristiques de concision

II. — CRITIQUE

1. — DÉFINITION DU CRITÈRE.

Critiquer, c'est faire l'examen des idées d'un texte et donner un avis sur chacune d'elles.

2. — DESCRIPTION DES ASPECTS.

Cette définition impose les deux aspects retenus :
 — Le nombre et la correction des idées reprises dans la critique;
 — Le type et la validité des arguments utilisés.

L'échelle est conçue de façon à attribuer davantage d'importance au type d'argument utilisé et à leur validité, qu'au nombre et à la correction des idées reprises. L'argumentation est, en effet, l'essentiel de la critique.

Comme pour le résumé, on n'évalue pas les aspects formels (orthographe, etc.). En outre, on n'a pas prévu d'échelle spéciale pour l'appréciation de l'enchaînement des arguments. Il sera jugé en même temps que la validité.

I. — Nombre et correction des idées reprises dans la critique

On n'évalue ici que les idées discutées par un argument valide au moins.

1. Seules les critiques portant sur les idées d'Alain seront prises en considération, et non celles qui sont relatives au style, au vocabulaire, au ton employés par l'auteur.
2. L'élève ne critique pas
 - si son travail reprend une idée du texte en marquant son accord ou son désaccord, sans le justifier d'aucune manière;
 - s'il se contente de paraphraser l'auteur sans apporter d'élément nouveau.
3. On ne retiendra dans l'évaluation que les idées assorties d'au moins un argument valide. On préconisera au paragraphe II les critères de validité de l'argumentation.

II. — Type et validité des arguments utilisés

Cette évaluation se fera en deux phases successives :

a) On donnera tout d'abord un avis sur la validité de l'argument présenté.

Il n'appartient pas au professeur de mettre en cause la validité de l'argument avancé par l'élève pour le simple fait qu'il ne partage pas son opinion. Les critères retenus seront d'ordre logique.

Est tenu pour valide un argument qui ne présente :

- ni erreur de raisonnement ⁽¹⁾,
 Par exemple, le recours à des arguments non pertinents (sans rapport avec les idées qu'ils sont censés appuyer).
 . pétitions de principe (reprennent, dans la démonstration, les arguments à démontrer).

⁽¹⁾ Cette liste n'est pas destinée à la découverte des types d'erreurs de raisonnement que contient le travail d'un élève. Elle n'est d'ailleurs pas exhaustive, mais aide le professeur à déterminer si un argument est valide ou non.

- . erreurs sur la cause (relation d'antériorité considérée comme relation causale).
- . généralisation hâtive (règle tirée d'un seul exemple).
- . faux dilemme (réduit en un dilemme une situation non ainsi réductible).

— ni contradiction interne.

Il y a contradiction interne quand l'argument entre en opposition avec un argument précédemment avancé par l'élève : *ce second argument est alors considéré comme non valide*. La contradiction peut exister entre deux arguments relatifs à la même idée ou à deux idées différentes.

Un argument non valide est annulé. Si une idée n'est appuyée que par un seul argument non valide, on considérera qu'elle n'est pas critiquée et on n'en tiendra pas compte pour l'évaluation du premier aspect.

b) *Si un argument est valide, on examinera à laquelle des trois catégories suivantes il appartient.*

Type A. Une argumentation subjective basée sur des jugements de valeur d'ordre moral ou sur des affirmations que l'élève n'étaye pas par des faits observables.

Type B. Une argumentation objective basée sur des faits observables précisés. A la différence des arguments du type A, on exige ici que l'élève indique les faits sur lesquels il s'appuie.

Type C. Une argumentation objective basée sur des implications ou des anticipations.

- . Implications : L'élève tire des idées d'Alain des conséquences qu'Alain n'a pas envisagées lui-même.

- . Anticipations : L'élève prévoit des objections qui pourraient lui être opposées et y répond.

3. — DÉTERMINATION DES NIVEAUX D'EXIGENCE.

a) Pour l'aspect « Nombre et correction des idées reprises dans la critique », on retiendra les niveaux suivants :

1. Critique d'au moins deux idées principales complètes.
2. Critique d'au moins une idée principale.
3. Critique d'une idée partielle au moins.
4. Critique d'une idée faussée au moins ⁽¹⁾.
5. Absence de critique.

b) Pour l'aspect « Type et validité des arguments utilisés », on distinguera aussi trois niveaux :

1. Argument(s) du type C (accompagnés éventuellement d'arguments du type A et/ou B).
2. Argument(s) du type B (accompagnés éventuellement d'arguments du type A).
3. Argument(s) du type A seulement.

⁽¹⁾ La spécificité de la performance évaluée par ce critère justifie ce quatrième niveau d'exigence qui permet de tenir compte de l'effort d'argumentation de l'élève.

Ces exigences doivent être retenues pour aucune des idées prises isolément. Il est cependant préférable de mentionner qu'il soutient.

4. — DÉMARCHES POUR L'ÉVALUATION

L'évaluation des idées reprises dans les arguments associés à ces idées (2^e aspect).

C'est pourquoi il importe, dans le tableau I, de distinguer les caractéristiques répertoriées dans le travail

1 ^{er} aspect : Idée critiquée		
Idée A — Idée complète critiquée <input type="checkbox"/>	non critiquée <input type="checkbox"/>	par
		non
— Idée partielle critiquée <input type="checkbox"/>	non critiquée <input type="checkbox"/>	par
		non
— Idée faussée critiquée <input type="checkbox"/>	non critiquée <input type="checkbox"/>	par
		non
— Idée omise <input type="checkbox"/>		

é considérée comme relation causale).
l exemple).
situation non ainsi réductible).

ment entre en opposition avec un argu-
ce second argument est alors considéré
exister entre deux arguments relatifs

l'idée n'est appuyée que par un seul
est pas critiquée et on n'en tiendra pas

laquelle des trois catégories suivantes il

é sur des jugements de valeur d'ordre
l'élève n'étaye pas par des faits obser-

é sur des faits observables précisés
de A, on exige ici que l'élève indique les

é sur des implications ou des anticipa-

ées d'Alain des conséquences qu'Alain

des objections qui pourraient lui être

es reprises dans la critique », on retien-

es complètes.

ats utilisés », on distinguera aussi trois

ventuellement d'arguments du type A

ventuellement d'arguments du type A).

e par ce critère justifie ce quatrième
de l'effort d'argumentation de l'élève.

Ces exigences doivent être remplies pour l'ensemble du travail et non pour cha-
cune des idées prises isolément. Pour les commentaires à communiquer à l'élève,
il est cependant préférable de mentionner chaque argument en relation avec l'idée
qu'il soutient.

4. — DÉMARCHE POUR L'ÉVALUATION DE LA CRITIQUE.

L'évaluation des idées reprises dans la critique (1^{er} aspect) dépend en partie des
arguments associés à ces idées (2^e aspect).

C'est pourquoi il importe, dans un premier temps, d'analyser chacune des idées
pour les deux aspects (tableau I). En un deuxième temps, on synthétisera les carac-
téristiques réperées dans le travail (tableau II).

1 ^{er} aspect : Idée critiquée	2 ^e aspect : Type et validité des arguments utilisés		L'idée est-elle reprise dans l'évaluation ?
	Validité de l'argument	Type d'argument	
Idée A — Idée complète critiquée <input type="checkbox"/>	par un argument <input type="checkbox"/> valide ou moins	A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/>	oui
	par un argument <input type="checkbox"/> non valide		non
	non critiquée <input type="checkbox"/>		non
— Idée partielle critiquée <input type="checkbox"/>	par un argument <input type="checkbox"/> valide ou moins	A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/>	oui
	par un argument <input type="checkbox"/> non valide		non
	non critiquée <input type="checkbox"/>		non
— Idée faussée critiquée <input type="checkbox"/>	par un argument <input type="checkbox"/> valide ou moins	A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/>	oui
	par un argument <input type="checkbox"/> non valide		non
	non critiquée <input type="checkbox"/>		non
— Idée omise <input type="checkbox"/>			non

1 ^{er} aspect : Idée critiquée	2 ^e aspect : Type et validité des arguments utilisés		L'idée est-elle reprise dans l'évaluation ?
	Validité de l'argument	Type d'argument	
Idée B — Idée complète critiquée <input type="checkbox"/> non critiquée <input type="checkbox"/>	par un argument <input type="checkbox"/> valide ou moins	A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/>	oui
	par un argument <input type="checkbox"/> non valide		non
			non
— Idée partielle critiquée <input type="checkbox"/> non critiquée <input type="checkbox"/>	par un argument <input type="checkbox"/> valide ou moins	A <input type="checkbox"/> B <input type="checkbox"/> c <input type="checkbox"/>	oui
	par un argument <input type="checkbox"/> non valide		non
			non
— Idée faussée critiquée <input type="checkbox"/> non critiquée <input type="checkbox"/>	par un argument <input type="checkbox"/> valide ou moins	A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/>	oui
	par un argument <input type="checkbox"/> non valide		non
			non
— Idée omise <input type="checkbox"/>			non

TABLEAU II

1. Faites porter votre appréciation sur le premier aspect : « Nombre et correction des idées ».

La critique contient :

- Au moins deux idées principales complètes et correctement comprises
- Au moins une idée principale et correctement comprise
- Au moins une idée partielle et correctement comprise
- Des idées faussées seulement
- Absence de critique

2. Faites porter votre appréciation sur les arguments utilisés ».

La critique contient :

- Un ou des arguments du type éventuellement d'arguments d
- Un ou des arguments du type éventuellement d'arguments d
- Un ou des arguments du type

3. En vous reportant au schéma su caractéristiques que vous venez d

Nombre et correction des idées discutées

- Au moins deux idées principales complètes correctement comprises

Au moins une idée principale complète et correctement comprise

- Au moins une idée partielle et correctement comprise

- Idées faussées seulement

- Absence de critique

Type et validité des arguments utilisés		L'idée est-elle reprise dans l'évaluation ?
Type d'argument		
ent <input type="checkbox"/>	A <input type="checkbox"/>	oui
ns <input type="checkbox"/>	B <input type="checkbox"/>	
	C <input type="checkbox"/>	
ent <input type="checkbox"/>		non
ent <input type="checkbox"/>		non
ent <input type="checkbox"/>	A <input type="checkbox"/>	oui
ns <input type="checkbox"/>	B <input type="checkbox"/>	
	c <input type="checkbox"/>	
ent <input type="checkbox"/>		non
ent <input type="checkbox"/>		non
ent <input type="checkbox"/>	A <input type="checkbox"/>	oui
ns <input type="checkbox"/>	B <input type="checkbox"/>	
	C <input type="checkbox"/>	
ent <input type="checkbox"/>		non
		non
		non

II

ier aspect : « Nombre et correction des idées discutées »

es et correctement comprises

ment comprise

ment comprise

2. Faites porter votre appréciation sur le deuxième aspect : « Type et validité des arguments utilisés ».

La critique contient :

- Un ou des arguments du type C (accompagnés éventuellement d'arguments de type B et/ou A)
- Un ou des arguments du type B (accompagnés éventuellement d'arguments du type A)
- Un ou des arguments du type A seulement

3. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux deux caractéristiques que vous venez d'évaluer dans le travail de l'élève.

Nombre et correction des idées discutées	Type et validité des arguments
— Au moins deux idées principales complètes correctement comprises	— Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A).... TB
	— Argument(s) du type B (accompagnés éventuellement d'arguments A) B
	— Argument(s) du type A seulement S
— Au moins une idée principale complète et correctement comprise	— Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A).... B
	— Argument(s) du type B (accompagnés éventuellement d'arguments A) S
	— Argument(s) du type A seulement F
— Au moins une idée partielle et correctement comprise	— Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A).... S
	— Argument(s) du type B (accompagnés éventuellement d'arguments A) F
	— Argument(s) du type A seulement I
— Idées faussées seulement	— Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A).... F
	— Argument(s) du type B (accompagnés éventuellement d'arguments A) I
	— Arguments du type A seulement I
— Absence de critique	I

Note pour la critique :

5. — PRÉSENTATION DE L'ÉCHELLE.

	Nombre et corrections des idées discutées	Type et validité des arguments
TB	Au moins deux idées principales complètes et correctement comprises	Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
B	Au moins deux idées principales complètes et correctement comprises	Argument(s) du type B (accompagnés éventuellement d'arguments A)
	Au moins une idée principale complète et correctement comprise	Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
S	Au moins deux idées principales complètes et correctement comprises	Argument(s) du type A seulement
	Au moins une idée principale complète et correctement comprise	Argument(s) du type B (accompagnés éventuellement d'arguments A)
	Au moins une idée partielle et correctement comprise	Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
F	Au moins une idée principale complète et correctement comprise	Argument(s) du type A seulement
	Au moins une idée partielle et correctement comprise	Argument(s) du type B (accompagnés éventuellement d'arguments A)
	Idées faussées seulement	Argument(s) du type C (accompagnés d'arguments B et/ou A)
I	Au moins une idée partielle et correctement comprise	Argument(s) du type A seulement
	Idées faussées seulement	Arguments du type B ou A
	Absence de critique	

La synthèse de plusieurs évaluations de performance est appréciée en fonction de la qualité de la synthèse, de la pertinence de la sélection, on synthétisera les évaluations retenues en fonction du choix du vocabulaire, etc. La note sera attribuée en fonction de l'importance respective accordée à ces différents critères.

Dans le cas du travail présenté dans les performances distinctes. Il semblerait que...

SYNTHÈSE DES ÉVALUATIONS

La synthèse de plusieurs évaluations analytiques se justifie quand une même performance est appréciée en fonction de différents critères. Par exemple, pour la rédaction, on synthétisera les évaluations relatives aux idées exprimées, à leur organisation, au choix du vocabulaire, etc. La note de synthèse apporte à l'élève des informations sur l'importance respective accordée par le professeur à chacun des critères.

Dans le cas du travail présenté dans ce dossier, le résumé et la critique constituent des performances distinctes. Il semble inopportun d'en synthétiser les évaluations.

Type et validité des arguments
Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
Argument(s) du type B (accompagnés éventuellement d'arguments A)
Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
Argument(s) du type A seulement
Argument(s) du type B (accompagnés éventuellement d'arguments A)
Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
Argument(s) du type A seulement
Argument(s) du type B (accompagnés éventuellement d'arguments A)
Argument(s) du type C (accompagnés éventuellement d'arguments B et/ou A)
Argument(s) du type A seulement
Arguments du type B ou A
critique

TRAVAUX SCIENTIFIQUES

Enseignement

ÉCHANGES

COMPTE RENDU D'UN
INTERPRÉTATION, INVARIANTE

Ce dossier est dû à la collaboration

M. DIGHAYE, Inspecteur de sciences

Mmes LANGE, E.N.M.E., Liège;
RUHWIEDEL, E.N.M.E., Liège

MM. AMEL, A.R. Liège II;
LHONEUX, A.R. Jupille;
PREGALDIEN, A.R. Visé

Chercheurs : J. PAQUAY-BECKERS et

TRAVAUX SCIENTIFIQUES EXPÉRIMENTAUX

Enseignement secondaire rénové : 1^{er} cycle

ÉCHELLES D'ÉVALUATION

COMPTE RENDU D'UNE EXPÉRIENCE SCIENTIFIQUE,
INTERPRÉTATION, INVENTION DE PROCÉDÉS DE MESURE

Ce dossier est dû à la collaboration d'une commission de professeurs composée de :

M. DIGHAYE, Inspecteur de sciences dans l'enseignement secondaire.

Mmes LANGE, E.N.M.E., Liège;
RUHWIEDEL, E.N.M.E., Liège.

MM. AMEL, A.R. Liège II;
LHONEUX, A.R. Jupille;
PREGALDIEN, A.R. Visé.

Chercheurs : J. PAQUAY-BECKERS et R. DE BAL.

L'évaluation fait partie intégrante du processus d'enseignement dont elle est une composante essentielle. Aussi importe-t-il de garantir la *validité* des instruments de mesure utilisés. Un instrument valide est représentatif de la matière dont on souhaite vérifier l'apprentissage : les questions explorent à la fois tous les contenus et tous les comportements applicables à ces contenus. Une telle procédure nécessite des études théoriques qui recensent les éléments fondamentaux de l'acquisition d'une notion. L'évaluation permet alors d'informer chaque élève de son progrès vers un objectif déterminé. En outre, la performance de l'élève peut être comparée à celles d'une population de référence (étalonnage normatif). Nous sommes encore loin de disposer des études et des instruments nécessaires...

Pour remplir son rôle vis-à-vis de l'élève, l'évaluation doit être aussi *objective* et *nuancée* que possible. Le degré de concordance entre différents juges est un indice important de l'objectivité d'une note. Or de nombreuses études docimologiques ont montré, pratiquement pour toutes les disciplines, combien les juges s'accordent peu sur la note à attribuer à une performance donnée. De plus, l'élève ne peut tirer un profit réel de l'évaluation que si les informations qui l'accompagnent sont analytiques, complètes et précises. Pour que le professeur puisse apporter un tel *feedback* à l'élève, la démarche qui préside à l'attribution de la note doit, elle aussi, être analytique.

Les difficultés de l'évaluation sont d'autant plus grandes que la performance sollicitée de l'élève nécessite la mise en œuvre de connaissances et d'aptitudes nombreuses. Les échelles descriptives présentées ici ont été construites pour l'évaluation de performances complexes. Des professeurs ont mis leur expérience en commun pour choisir des critères d'appréciation aussi sûrs, aussi valides que possible. Ces critères ont été appliqués à de nombreux travaux.

Dans l'espoir d'assurer l'homogénéité des notes, des consignes de correction détaillées sont proposées (et justifiées).

L'évaluation à l'aide d'échelles descriptives fournit deux types d'informations à l'élève :

- Il est informé des points forts et des lacunes de son travail;
- Il dispose de points de repère pour situer sa performance par rapport à ses condisciples.

En raison de son caractère analytique, l'évaluation à l'aide d'échelles descriptives prend beaucoup de temps : elle ne pourrait pas être pratiquée quotidiennement dans une classe. Elle devrait être réservée pour certains moments cruciaux de l'apprentissage, ou, plus simplement, pour des vérifications périodiques. Au lieu de se servir des batteries complètes qui lui sont proposées, le professeur a aussi toujours le loisir de n'utiliser qu'une ou deux des échelles, selon les besoins.

On peut espérer que, petit à petit, se créeront des banques d'instruments d'évaluation auxquelles les enseignants pourront faire appel à leur convenance et à celle de leurs élèves.

Notes pour l'évaluation des travaux.

1. Les critères d'évaluation proposés dans les échelles ci-après ont été définis et testés par une commission de professeurs, au cours de nombreuses séances de travail. Il se peut que ces critères ne soient pas ceux en fonction desquels vous jugez habituellement. Pour essayer l'instrument, il importe cependant que vous vous conformiez strictement aux consignes de correction proposées.

En vue de modifications ultérieures seront précieuses.

2. Pour la commodité et l'efficacité de vos travaux de vos élèves selon la première suivantes.
3. En guise d'entraînement, un travail

cessus d'enseignement dont elle est une
e garantir la *validité* des instruments de
présentatif de la matière dont on souhaite
rent à la fois tous les contenus et tous
nus. Une telle procédure nécessite des
ts fondamentaux de l'acquisition d'une
er chaque élève de son progrès vers un
e de l'élève peut être comparée à celles
normatif). Nous sommes encore loin de
aires...

l'évaluation doit être aussi *objective* et
ance entre différents juges est un indice
de nombreuses études docimologiques
disciplines, combien les juges s'accordent
ce donnée. De plus, l'élève ne peut tirer
rmations qui l'accompagnent sont ana-
fesseur puisse apporter un tel *feedback* à
de la note doit, elle aussi, être analytique.

t plus grandes que la performance solli-
connaissances et d'aptitudes nombreuses.
té construites pour l'évaluation de per-
mis leur expérience en commun pour
aussi valides que possible. Ces critères

es notes, des consignes de correction

es fournit deux types d'informations à

unes de son travail;

sa performance par rapport à ses condis-

valuation à l'aide d'échelles descriptives
pas être pratiquée quotidiennement dans
ins moments cruciaux de l'apprentissage,
ériodiques. Au lieu de se servir des bat-
fesseur a aussi toujours le loisir de n'utili-
soins.

t des banques d'instruments d'évaluation
el à leur convenance et à celle de leurs

échelles ci-après ont été définis et testés
urs de nombreuses séances de travail. Il
x en fonction desquels vous jugez habi-
mporte cependant que vous vous confor-
tion proposées.

En vue de modifications ultérieures des critères, vos remarques et vos critiques
seront précieuses.

2. Pour la commodité et l'efficacité de l'évaluation, il est conseillé d'évaluer tous les
travaux de vos élèves selon la première échelle, puis de procéder de même pour les
suivantes.
3. En guise d'entraînement, un travail corrigé est proposé en annexe.

A. — Description de l'exercice choisi

1. — OBJECTIFS.

En début d'année, des élèves de première année de l'enseignement secondaire rénové ont été invités à :

- observer une expérience relative à la dureté des eaux;
- interpréter les phénomènes observés;
- imaginer une méthode de mesure.

2. — PRÉSENTATION DE L'EXPÉRIENCE.

a) Introduction.

On rappelle d'abord quelques observations de la vie courante qui permettraient de répondre à la question : « Utilise-t-on n'importe quelle eau pour n'importe quel usage ? ».

b) Expérience.

Le professeur réalise une expérience devant tous les élèves en utilisant de grands récipients. Les seules informations fournies aux élèves ne concernent que la nature des produits utilisés.

A. Matériel.

- . Trois éprouvettes graduées (A, B, C) d'un demi-litre, contenant chacune 490 cc d'eau distillée.
- . Un récipient contenant du sel soluble de calcium.
- . Une solution alcoolique de savon (Boutron-Boudet).
- . Une mesurette pour le sel de calcium.
- . Une mesurette pour la solution de savon.

B. Déroulement de l'expérience.

- . Jeter une ration de sel soluble de calcium dans l'éprouvette B.
- . Jeter deux rations de sel de calcium dans l'éprouvette C.
- . Ajouter une ration de solution de savon dans l'éprouvette A : la mousse apparaît immédiatement après agitation.
- . Ajouter la même quantité de savon dans les éprouvettes B et C : la mousse n'apparaît pas.
- . Continuer à verser, à la fois dans B et C, des rations de solution de savon jusqu'à apparition de mousse persistante dans chacune des éprouvettes et la formation d'un précipité.

C. Tâches demandées aux élèves.

1. Décrivez les phénomènes que vous venez d'observer.
2. Relatez l'expérience par des croquis.
3. Expliquez les phénomènes observés.
4. Inventez un ou plusieurs procédés de mesure de la quantité de sel de calcium que contiennent différentes eaux.

B. — Justification

Cet exercice demande une performance de déroulement de l'expérience.

L'évaluation ne peut se limiter à un résultat, elle nécessite la mise au point d'un instrument.

C. — Choix des

Le plan de l'exercice proposé à l'élève

1. Relation de l'expérience par une description.
2. Relation de l'expérience par des schémas.
3. Interprétation des phénomènes observés.
4. Invention de procédés de mesure.

Ces dimensions s'appliquent à des

D. — Or

Pour chacune des dimensions, on

- La définition du critère.
- La définition des aspects qui interviennent, toujours un poids égal dans l'évaluation.
- La détermination des niveaux d'exigence. Les exigences, pour chaque dimension, ont été établies après un examen de travaux d'élèves de première année.
- La démarche de l'évaluation.
- La présentation de l'échelle.

Cette présentation n'apporte pas de nouveauté à la démarche. Toutefois, elle établit une hiérarchie de la performance.

ière année de l'enseignement secondaire
été des eaux;

ons de la vie courante qui permettraient
n'importe quelle eau pour n'importe quel

ant tous les élèves en utilisant de grands
aux élèves ne concernent que la nautre

un demi-litre, contenant chacune 490 cc
calcium.
(Bouquet).

ans l'éprouvette B,
l'éprouvette C.
ans l'éprouvette A : la mousse apparaît
es éprouvettes B et C : la mousse n'appa-

des rations de solution de savon jusqu'à
cune des éprouvettes et la formation d'un

d'observer.

sure de la quantité de sel de calcium que

B. — Justification du choix de cet exercice

Cet exercice demande une performance complexe aux élèves qui ont assisté au déroulement de l'expérience.

L'évaluation ne peut se limiter à un simple décompte de points par faute commise; elle nécessite la mise au point d'un instrument plus adéquat.

C. — Choix des dimensions de l'évaluation

Le plan de l'exercice proposé à l'élève indique les quatre critères retenus :

1. Relation de l'expérience par une description écrite;
2. Relation de l'expérience par des schémas;
3. Interprétation des phénomènes observés;
4. Invention de procédés de mesure.

Ces dimensions s'appliquent à de nombreux exercices scientifiques.

D. — Organisation du dossier

Pour chacune des dimensions, on présente successivement :

— *La définition du critère.*

— *La définition des aspects* qui interviennent dans chaque critère. Ces aspects auront toujours un poids égal dans l'évaluation.

— *La détermination des niveaux d'exigence.*

Les exigences, pour chaque degré d'une échelle, ne sont pas arbitraires : elles ont été établies après une analyse approfondie d'un nombre important de travaux d'élèves de première année de l'enseignement secondaire rénové.

— *La démarche de l'évaluation.*

— *La présentation de l'échelle.*

Cette présentation n'apporte aucune information complémentaire par rapport à la démarche. Toutefois, elle offre l'avantage de donner une description hiérarchique de la performance.

PRÉSENTATION DES ÉCHELLES

I. — RELATION DE L'EXPÉRIENCE PAR UNE DESCRIPTION ÉCRITE

A. — Définition du critère

Traduction écrite des phénomènes observés.

B. — Définition des aspects qui interviennent dans le critère

On évalue successivement trois aspects : le caractère *complet* du travail, la *précision* des informations et le respect de la *chronologie*.

A. Un travail *complet* devra comporter les informations suivantes :

- A1. *Egalité* des volumes d'eau distillée de départ ⁽¹⁾.
- A2. *Inégalité* des quantités de sel de calcium ajoutés aux deux volumes d'eau.
- A3. *Différence de limpidité* entre les contenus des trois éprouvettes.
- A4. *Inégalité des doses de savon* conduisant à l'apparition d'une mousse persistante.

B. Un travail *précis* contiendra les informations suivantes :

- B1. Valeurs relatives des quantités de sels de Ca.
(Ex. : simple - double, 1 dose - 2 doses, 1 g - 2 g).
- B2. Valeurs relatives des doses de savon.
(Ex. : 6 mesures - 13 mesures).

C. La suite chronologique des phénomènes est respectée quand l'élève rend compte des faits dans l'ordre où ils se sont déroulés.

On exigera, au minimum, que l'élève relate l'adjonction du sel de calcium avant l'adjonction du savon.

C. — Détermination des niveaux d'exigence

a) Pour l'aspect « Caractère complet », on retiendra trois niveaux :

- 1. Les informations A1, A2, A3 et A4 sont présentes;
- 2. Les informations A2 et A4 au moins, sont présentes;
- 3. L'information A2 ou l'information A4 manque.

b) Pour l'aspect « Précision », on retiendra trois niveaux :

- 1. Les informations B1 et B2 sont présentes;
- 2. L'information B1 ou l'information B2 est présente;
- 3. Les informations B1 et B2 manquent.

⁽¹⁾ La caractéristique A1 est présente si l'élève mentionne à la fois qu'il s'agit d'eau distillée et que les volumes sont égaux. Si l'une de ces informations manque ou est entachée d'erreur, le travail ne comporte pas la caractéristique A1.

c) Pour l'aspect « Respect de la chr

- 1. Oui;
- 2. Non.

D. — Démarche pour l'évaluat

1. Appréciez le premier aspect : « C

- A1, A2, A3 et A4 ⁽¹⁾
- A2 et A4
- A2 ou A4 manque

2. Appréciez le deuxième aspect : «

- B1 et B2
- B1 ou B2
- ni B1 ni B2

3. Appréciez le troisième aspect : «

- Oui
- Non

4. En vous reportant au schéma su
caractéristiques que vous venez

Caractère complet *Préc*

— A1, A2, A3 et A4

— A2 et A4

— A2 ou A4 manque

⁽¹⁾ Indiquez par une croix, dan
choisissez.

ÉCHELLES

UNE DESCRIPTION ÉCRITE

u critère

erviennent dans le critère

ctère complet du travail, la précision

mations suivantes :

part (1).
m ajoutés aux deux volumes d'eau.
us des trois éprouvettes.
t à l'apparition d'une mousse per-

suivantes :

e Ca.
1 g - 2 g).

respectée quand l'élève rend compte

late l'adjonction du sel de calcium

veaux d'exigence

ndra trois niveaux :

présentes;
présentes;
nque.

niveaux :

présente;

ève mentionne à la fois qu'il s'agit
si l'une de ces informations manque
pas la caractéristique A1.

e) Pour l'aspect « Respect de la chronologie », on retiendra deux niveaux :

1. Oui;
2. Non.

D. — Démarche pour l'évaluation de la « Description écrite de l'expérience »

1. Appréciez le premier aspect : « Caractère complet » :

- A1, A2, A3 et A4 (1)
- A2 et A4
- A2 ou A4 manque

2. Appréciez le deuxième aspect : « Précision » :

- B1 et B2
- B1 ou B2
- ni B1 ni B2

3. Appréciez le troisième aspect : « Respect de la chronologie ».

- Oui
- Non

4. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux trois caractéristiques que vous venez d'évaluer dans le travail de l'élève.

<i>Caractère complet</i>	<i>Précision</i>	<i>Respect de la chronologie</i>	
— A1, A2, A3 et A4	— B1 et B2	— Oui	TB
		— Non	B
	— B1 ou B2	— Oui	B
		— Non	S
	— ni B1 ni B2	— Oui	S
		— Non	F
— A2 et A4	— B1 et B2	— Oui	B
		— Non	S
	— B1 ou B2	— Oui	S
		— Non	F
	— ni B1 ni B2	— Oui	F
		— Non	I
— A2 ou A4 manque	— Quelles que soient les caractéristiques de précision et de chronologie		I

Note pour la description :

(1) Indiquez par une croix, dans la case prévue à cet effet, la solution que vous choisissez.

E. — Présentation de l'échelle : « Description écrite de l'expérience »

	Caractère complet	Précision	Respect de la chronologie
TB	A1, A2, A3 et A4	B1 et B2	Oui
B	A1, A2, A3 et A4	B1 et B2	Non
	A1, A2, A3 et A4	B1 et B2	Oui
	A2 et A4	B1 et B2	Oui
S	A1, A2, A3 et A4	B1 ou B2	Non
	A1, A2, A3 et A4	ni B1 ni B2	Oui
	A2 et A4	B1 et B2	Non
	A2 et A4	B1 ou B2	Oui
F	A1, A2, A3 et A4	ni B1 ni B2	Non
	A2 et A4	ni B1 ni B2	Non
	A2 et A4	ni B1 ni B2	Oui
I	A2 et A4	ni B1 ni B2	Non
	A2 ou A4 manque	Quelles que soient les caractéristiques de précision et de chronologie	

II. — RELATION DE L'É

A. — De

On évalue ici à la fois les schémas complètent.

N.B. : L'annotation d'un élément ne doit pas être nécessairement répétée

B. — Définition des aspects

On évalue successivement trois aspects : le caractère *précis* et le caractère *dynamique*

1. Caractère complet.

Pour que le schéma de l'élève soit en forme de dessins ou d'annotations, un schéma doit comporter :
 A1. Au départ de l'expérience, égalité de volumes.
 A2. Addition de doses inégales de sel.
 A3. Addition de doses inégales de savon.
 A4. Apparition de mousse dans les tubes.
 A5. Différence d'opacité de la solution.

2. Caractère précis.

Les informations A2 et A3 peuvent être précises en spécifiant les valeurs relatives...

B1. des quantités de sel de calcium

B2. des quantités de savons (6 doses)

Une information importante — la présence de mousse — n'entre pas dans l'évaluation de l'échantillon.

3. Caractère dynamique.

Les croquis de l'élève présentent le déroulement des étapes de l'expérience.

Le déroulement complet de l'expérience est :

1. Trois éprouvettes identiques contenant du liquide.
2. Les trois éprouvettes avec adjonction de sel et C.

(²) Certains élèves utilisent le mot "mousse" si l'information A4 est présente, mais

Précision	Respect de la chronologie
et B2	Oui
et B2	Non
et B2	Oui
et B2	Oui
ou B2	Non
1 ni B2	Oui
et B2	Non
ou B2	Oui
1 ni B2	Non
1 ni B2	Non
1 ni B2	Oui
1 ni B2	Non
ne soient les caractéristiques de la description et de chronologie	

II. — RELATION DE L'EXPÉRIENCE À L'AIDE DE SCHÉMAS

A. — Définition du critère

On évalue ici à la fois les schémas réalisés par l'élève et les annotations qui les complètent.

N.B. : L'annotation d'un élément identique pour différentes parties du schéma ne doit pas être nécessairement répétée.

B. — Définition des aspects qui interviennent dans le critère

On évalue successivement trois aspects du schéma de l'élève : le caractère *complet*, le caractère *précis* et le caractère *dynamique*.

1. Caractère complet.

Pour que le schéma de l'élève soit considéré comme complet, il doit fournir, sous forme de dessins ou d'annotations, un certain nombre d'informations :

- A1. Au départ de l'expérience, égalité d'eau distillée dans les trois éprouvettes.
- A2. Addition de doses inégales de sels de calcium dans les éprouvettes B et C.
- A3. Addition de doses inégales de savon dans les trois éprouvettes.
- A4. Apparition de mousse dans les trois éprouvettes après addition de la dose nécessaire de savon ⁽¹⁾.
- A5. Différence d'opacité de la solution dans les éprouvettes B et C.

2. Caractère précis.

Les informations A2 et A3 peuvent être complétées par des précisions quantitatives spécifiant les valeurs relatives...

- B1. des quantités de sel de calcium (1 dose - 2 doses);
- B2. des quantités de savons (6 doses - 13 doses).

Une information importante — la nécessité de représenter trois éprouvettes identiques — n'entre pas dans l'évaluation : elle est présente dans toutes les copies de l'échantillon.

3. Caractère dynamique.

Les croquis de l'élève présentent cette caractéristique s'ils permettent de cerner le déroulement des étapes de l'expérience, par plusieurs séries de dessins.

Le déroulement complet de l'expérience comporte les étapes suivantes :

1. Trois éprouvettes identiques contenant un volume égal d'eau distillée (état initial).
2. Les trois éprouvettes avec adjonction de sel de calcium dans les éprouvettes B et C.

⁽¹⁾ Certains élèves utilisent le mot « savon » pour désigner la mousse. Dans ce cas, l'information A4 est présente, mais l'information A3 ne l'est pas.

3. Les trois éprouvettes avec adjonction de savon et apparition de mousse (pour cette phase de l'expérience, l'élève pourrait montrer qu'après l'addition de chaque dose de savon, on agite l'éprouvette pour observer l'apparition éventuelle de mousse persistante).
4. Trois éprouvettes avec différence d'opacité (état final).
Au minimum, l'élève doit mentionner les points 2, 3 et 4. On acceptera le regroupement des informations relatives aux points 2 et 3, ou 3 et 4, en une même série de dessins.

C. — Détermination des niveaux d'exigence

- a) Pour l'aspect « Caractère complet », on retiendra trois niveaux :
 1. Les informations A1, A2, A3, A4 et A5 sont présentes.
 2. Les informations A2, A3 et A4 au moins sont présentes.
 3. Au moins une des informations A2, A3 ou A4 manque.
- b) Pour l'aspect « Caractère précis », on retiendra deux niveaux :
 1. Les informations B1 et B2 sont présentes.
 2. L'information B1 ou B2 manque.
- c) Pour l'aspect « Caractère dynamique », on retiendra deux niveaux :
 1. Oui;
 2. Non.

D. — Démarche pour l'évaluation de la « Relation de l'expérience par un schéma »

1. Appréciez le premier aspect : « Caractère complet ».
 - Toutes les informations A
 - A2, A3 et A4 au moins
 - A2, A3 ou A4 manque
2. Apprécier le second aspect : « Caractère précis ».
 - B1 et B2
 - B1 ou B2 manque
3. Appréciez le troisième aspect : « Caractère dynamique ».
 - Oui
 - Non

4. En vous reportant au schéma s caractéristiques que vous venez de définir, indiquez le caractère de votre expérience.

Caractère complet

— Toutes les informations A

— A2, A3 et A4 au moins

— A2, A3 ou A4 manque ...

E. — Présentation de l'échelle

	Caractère complet
TB	Toutes les informations A
B	Toutes les informations A2, A3 et A4 au moins
	Toutes les informations A2, A3 ou A4 manquent
S	Toutes les informations B1 et B2
	Au moins une des informations B1 ou B2 manque
F	Oui
I	Non

4. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux trois caractéristiques que vous venez d'évaluer dans le travail de l'élève.

<i>Caractère complet</i>	<i>Caractère précis</i>	<i>Caractère dynamique</i>
— Toutes les informations A	{ — B1 et B2 — B1 ou B2 manque	{ — Oui TB
		{ — Non B
— A2, A3 et A4 au moins	{ — B1 et B2 — B1 ou B2 manque	{ — Oui B
		{ — Non S
— A2, A3 ou A4 manque		{ — Oui S
		{ — Non F
		I

Note pour le schéma :

E. — *Présentation de l'échelle : « Relation de l'expérience par un schéma »*

	Caractère complet	Caractère précis	Caractère dynamique
TB	Toutes les informations A	Informations B1 et B2	Oui
B	Toutes les informations A	B1 et B2	Non
	Toutes les informations A	B1 ou B2 manque	Oui
	A2, A3 et A4 au moins	B1 et B2	Oui
S	Toutes les informations A	B1 ou B2 manque	Non
	A2, A3 et A4 au moins	B1 et B2	Non
	A2, A3 et A4 au moins	B1 ou B2 manque	Oui
F	A2, A3 et A4 au moins	B1 ou B2 manque	Non
I	Au moins une des informations A2, A3 ou A4 manque	Quelles que soient les appréciations portées pour les caractères précis et dynamique	

III. — INTERPRÉTATION DE L'EXPÉRIENCE

A. — Définition du critère

On évalue les commentaires de l'élève qui, au-delà de la description des étapes de l'expérience, ont pour but de mettre en évidence les *relations* entre les phénomènes observés pour leur donner une *signification*.

B. — Définition des aspects qui interviennent dans le critère

On évalue successivement deux aspects : le caractère complet et le caractère précis.

1° Le caractère complet.

L'interprétation de l'élève est complète si elle met en rapport :

- A1. La présence de sel de calcium et la quantité de savon nécessaires à l'apparition de mousse;
- A2. La présence de sel de calcium et la formation d'un précipité.

2° Le caractère précis.

a) Les phénomènes observés peuvent être simplement mis en *relation* :

- La présence d'un sel de Ca impose l'addition d'une plus grande quantité de savon;
- La présence d'un sel de Ca entraîne la formation d'un précipité.

Sans nécessairement utiliser l'une de ces formulations, l'élève se contente, dans ce cas, d'exprimer un lien entre la présence des deux phénomènes.

b) Cette relation peut être précisée sous forme de *fonction* :

- Plus il y a de sel de Ca, plus il faut de savon pour obtenir de la mousse;
- Plus il y a de sel de Ca, plus le précipité est important.

Par rapport à la situation précédente, l'élève met en relation des *degrés* dans l'intensité des phénomènes observés.

c) La première de ces relations (sels de Ca - savon) peut prendre la forme d'une *proportion* :

- Pour 1 ration de sel de Ca, il faut 6 mesures de savon;
- Pour 2 rations de sel de Ca, il faut 12 mesures de savon.

Les quantités de sels de calcium et les quantités de savon varient dans le rapport de 1 à 6.

Ici, les degrés sont quantifiés par l'intervention d'unités de mesure qui permettent à l'élève d'exprimer le *rapport de grandeur* entre les quantités de sel de calcium et de savon.

N.B. — Les conditions de l'expérience, basée sur un seul exemple, rendent ce type de réponse peu probable.

L'analyse d'un grand nombre de copies montre d'ailleurs que cette formulation est rare en première année de l'enseignement secondaire rénové. Elle n'inter-

viendra donc pas dans l'échelle. Le prof éventuelle de cette formulation par un co

Dans l'échelle, cette performance sera exprimées sous forme de fonction au mo

C. — Détermination

Les deux aspects présentés plus haut s niveaux d'exigence comme si ces deux asp à suivre est donc simplifiée. Les niveau l'échelle.

D. — Présentation de l'échelle

TB	A1 et A2 s
B	A1 sous fo A2 sous fo
S	A1 et A2 s
	A1 sous fo
F	A2 sous fo
	A1 sous fo
I	Aucune in

N.B. — Pour les deux premiers cr démarche. Pour conserver à l'évaluatio lignera dans l'échelle la caractéristique

du critère

au-delà de la description des étapes
ence les *relations* entre les phénomènes

intervient dans le critère

caractère complet et le caractère précis.

met en rapport :

é de savon nécessaires à l'apparition de

tion d'un précipité.

plement mis en *relation* :

ddition d'une plus grande quantité de

ormation d'un précipité.

ormulations, l'élève se contente, dans
des deux phénomènes.

de *fonction* :

von pour obtenir de la mousse;

est important.

ève met en relation des *degrés* dans

savon) peut prendre la forme d'une

ures de savon;

esures de savon.

tités de savon varient dans le rapport

on d'unités de mesure qui per mettent à
re les quantités de sel de calcium et de

sur un seul exemple, rendent ce type

ontre d'ailleurs que cette formulation
ent secondaire rénové. Elle n'inter-

viendra donc pas dans l'échelle. Le professeur peut toutefois valoriser la présence éventuelle de cette formulation par un commentaire approprié.

Dans l'échelle, cette performance sera classée dans la catégorie « Informations exprimées sous forme de fonction au moins ».

C. — Détermination des niveaux d'exigence

Les deux aspects présentés plus haut sont indissociables. Aussi, on présentera les niveaux d'exigence comme si ces deux aspects n'en formaient qu'un seul. La démarche à suivre est donc simplifiée. Les niveaux d'exigence apparaissent clairement dans l'échelle.

D. — Présentation de l'échelle : « Interprétation de l'expérience »

TB	A1 et A2 sous forme de fonction au moins
B	A1 sous forme de fonction au moins et A2 sous forme de relation
S	A1 et A2 sous forme de relation
	A1 sous forme de fonction au moins
F	A2 sous forme de relation au moins
	A1 sous forme de relation
I	A2 sous forme de relation
	Aucune interprétation

N.B. — Pour les deux premiers critères, une page spéciale était réservée à la démarche. Pour conserver à l'évaluation un caractère analytique, le professeur soulignera dans l'échelle la caractéristique qui correspond au travail évalué.

Note pour l'interprétation :

A. — Définition du critère

La démarche d'invention suppose, au-delà de l'observation et de l'interprétation de phénomènes, la sélection des variables pertinentes permettant la mise en place d'un schéma général.

L'invention d'une méthode de mesure suppose, préalablement, une *interprétation correcte* de l'expérience observée. Pour qu'une réelle mesure soit possible, il faudrait que l'élève ait dégagé de l'observation la *proportion qui lie quantité de sel et quantité de savon*.

Toutefois, nous avons déjà souligné que l'expérience, telle qu'elle a été menée, ne permet pas de déterminer cette proportion. On se contentera donc du principe général : de la quantité de savon nécessaire pour faire mousser une eau, on peut déduire la quantité de sel de Ca contenue dans cette eau.

L'application de ce principe suppose le choix d'une *unité de mesure* et du procédé d'utilisation de celle-ci.

La seule méthode de *mesure* possible est donc fondée sur la relation entre savon et sel de calcium. Par contre, l'évaluation de l'opacité du précipité ne peut consister qu'en une *estimation*.

La méthode proposant l'évaporation du liquide et l'évaluation du résidu de sel de calcium subsistant au fond de l'éprouvette est bonne, mais *inappropriée au contexte de l'expérience*.

Si l'élève propose plusieurs procédés, on ne retiendra que le plus élaborée.

B. — Présentation de l'échelle : « Invention d'une méthode de mesure »

L'évaluation de cette dimension n'impose pas le recours à plusieurs aspects complémentaires. La démarche à suivre est donc simplifiée. Les niveaux d'exigence apparaissent directement dans l'échelle.

TB	Découverte : — du principe général (déduction de la quantité de sel à partir de la quantité de savon) — de la nécessité d'une unité de mesure — du procédé d'utilisation de cette unité
B	Découverte : — du principe général — de la nécessité d'une unité de mesure
S	Découverte du principe général Estimation de la quantité de sel de calcium par l'évaluation de l'opacité du précipité
F	Estimation du résidu de calcium après évaporation du liquide
I	Répétition de l'interprétation Absence de réponse

N.B. — Comme pour l'interprétation, souligner dans l'échelle, les caractéristiques qui correspondent au travail évalué.

Note pour l'invention d'une méthode de mesure :

Vous venez d'évaluer chacune des

Reportez les notes attribuées sur

1. Relation de l'expérience par une de

2. Relation de l'expérience à l'aide de

3. Interprétation des phénomènes obs

4. Invention de procédés de mesure

Note de synthèse

du critère

à de l'observation et de l'interprétation pertinentes permettant la mise en place

ppose, préalablement, une *interprétation* ne réelle mesure soit possible, il faudrait *proportion qui lie quantité de sel et quantité*

l'expérience, telle qu'elle a été menée, ne se contentera donc du principe général : e mousser une eau, on peut déduire la u.

oix d'une *unité de mesure* et du procédé

onc fondée sur la relation entre savon et opacité du précipité ne peut consister

uide et l'évaluation du résidu de sel de st bonne, mais *inappropriée au contexte*

e retiendra que le plus élaborée.

ention d'une *méthode de mesure* »

us le recours à plusieurs aspects complé- blifiée. Les niveaux d'exigence apparais-

duction de la quantité de sel à partir
ité de mesure de cette unité
ité de mesure
éral
de sel de calcium par l'évaluation
cium après évaporation du liquide
n

mer dans l'échelle, les caractéristiques

ention d'une méthode de mesure :

Vous venez d'évaluer chacune des dimensions.

Reportez les notes attribuées sur les échelles ci-après :

1. Relation de l'expérience par une description écrite

TB	B	S	F	I
 2. Relation de l'expérience à l'aide de schémas

TB	B	S	F	I
 3. Interprétation des phénomènes observés

TB	B	S	F	I
 4. Invention de procédés de mesure

TB	B	S	F	I
- Note de synthèse

TB	B	S	F	I

Elisa C.
1^{re} A.

Sciences - Rapport d'expérience

1 Il y a 3 éprouvettes sur la table, on y verse une quantité égale d'eau distillée. Après ceci dans la deuxième éprouvette on y a versé une petite dose de sel de calcium et dans la troisième 2 doses de sel de calcium. Les trois éprouvettes contiennent toujours 40^o cc d'eau distillée.

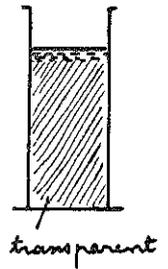
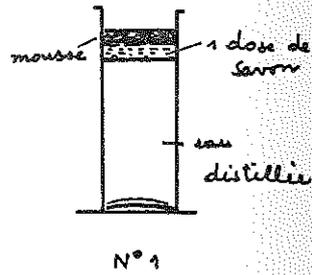
La première éprouvette toujours pleine d'eau pure contient maintenant 1 quantité de 10 cc de savon liquide. La même quantité est versée 5 fois dans la deuxième et 10 fois dans la troisième.

Ce mélange donne le résultat suivant :

La première éprouvette est toujours transparente et au-dessus il y a de la mousse car le savon a été mélangé dans les 3 cas. La deuxième éprouvette graduée est de couleur blanchâtre, mais pas transparente et au-dessus du mélange il y a aussi une petite hauteur de mousse le volume est plus élevé que dans la première éprouvette car la dose de savon est plus importante (60 cc)

La troisième éprouvette est beaucoup plus blanche que la 2^{me} et aussi de plus grand volume. La mousse est moins importante.

2



3

On peut aussi l'a
parce qu'elle n'en
se mélange et fait
autres éprouvettes qu
moins
plus abstr

4

On ajoute du savon de
mousse apparaît, i
elle n'apparaît pas

d'expérience

Table, on y verse une
après ceci dans la deuxième
dose de sel de calcium et dans
... Les trois éprouvette
distillée.

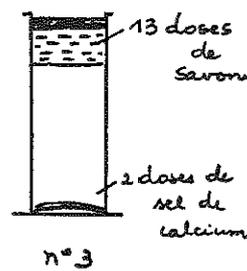
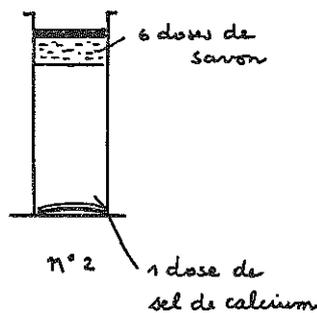
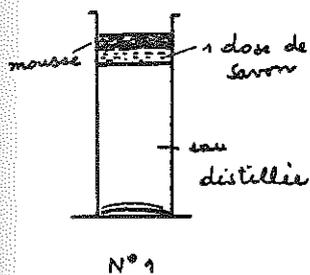
ins d'eau pure contient
de savon liquide. Le même
deuxième et 13 fois dans la

suivent :

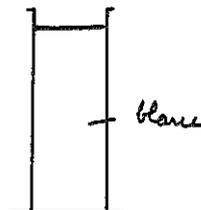
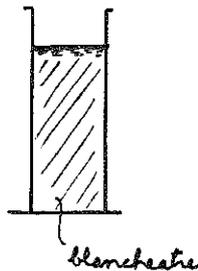
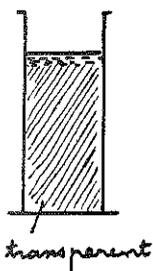
eurs transparents et au-
le savon a été mélangé
éprouvette graduée est de
transparente et au dessus
tite hauteur de mousse la
a première éprouvette car
tante (60cc)

beaucoup plus blanche que le
volume. la mousse est

2



après avoir mélangé le contenu
Voici la couleur obtenue



3

On peut aussi l'apercevoir dans l'eau distillée
parce qu'elle n'en contient pas c'est pour ça que le savon
se mélange et fait de la mousse tandis que dans les
autres éprouvettes qui contiennent du sel la mousse monte
^{moins}
~~plus vite~~

4

On ajoute du savon dans les différentes eaux. On agite. Si la
mousse apparaît, c'est qu'il n'y a pas de sel dans l'eau et si
elle n'apparaît pas c'est qu'il y a du sel.

SC

Enseignement sec

ÉCHELLES

OBSERVATION D'UNE EX
INVENTION D'UNE

Ce dossier est dû à la collaboration d

M. DIGHAYE, Inspecteur de scien

Mmes AMEL, A.R. de Montegnée;
LANGE, E.N.M.E. Liège.

MM. BOUCHAT, A.R. Jambes;
DEDOYARD, A.R. Angleur;
GUEIBE, A.R. Arlon;
OTTE, A.R. Liège II;
PIROTTE, A.R. Liège I.

Chercheurs : J. PAQUAY-BECKERS, R. I

SCIENCES

Enseignement secondaire rénové : 2^e cycle

ÉCHELLES D'ÉVALUATION

OBSERVATION D'UNE EXPÉRIENCE, INTERPRÉTATION, INVENTION D'UNE MÉTHODE D'ANALYSE

Ce dossier est dû à la collaboration d'une commission de professeurs composée de :

M. DIGHAYE, Inspecteur de sciences dans l'enseignement secondaire.

Mmes AMEL, A.R. de Montegnée;
LANGE, E.N.M.E. Liège.

MM. BOUCHAT, A.R. Jambes;
DEDOYARD, A.R. Angleur;
GUEIBE, A.R. Arlon;
OTTE, A.R. Liège II;
PIROTTE, A.R. Liège I.

Chercheurs : J. PAQUAY-BECKERS, R. DE BAL.

L'évaluation fait partie intégrante du processus d'enseignement dont elle est une composante essentielle. Aussi importe-t-il de garantir la *validité* des instruments de mesure utilisés. Un instrument valide est représentatif de la matière dont on souhaite vérifier l'apprentissage : les questions explorent à la fois tous les contenus et tous les comportements applicables à ces contenus. Une telle procédure nécessite des études théoriques qui recensent les éléments fondamentaux de l'acquisition d'une notion. L'évaluation permet alors d'informer chaque élève de son progrès vers un objectif déterminé. En outre, la performance de l'élève peut être comparée à celles d'une population de référence (étalonnage normatif). Nous sommes encore loin de disposer des études et des instruments nécessaires...

Pour remplir son rôle vis-à-vis de l'élève, l'évaluation doit être aussi *objective* et *nuancée* que possible. Le degré de concordance entre différents juges est un indice important de l'objectivité d'une note. Or de nombreuses études docimologiques ont montré, pratiquement pour toutes les disciplines, combien les juges s'accordent peu sur la note à attribuer à une performance donnée. De plus, l'élève ne peut tirer un profit réel de l'évaluation que si les informations qui l'accompagnent sont analytiques, complètes et précises. Pour que le professeur puisse apporter un tel *feedback* à l'élève, la démarche qui préside à l'attribution de la note doit, elle aussi, être analytique.

Les difficultés de l'évaluation sont d'autant plus grandes que la performance sollicitée de l'élève nécessite la mise en œuvre de connaissances et d'aptitudes nombreuses. Les échelles descriptives présentées ici ont été construites pour l'évaluation de performances complexes. Des professeurs ont mis leur expérience en commun pour choisir des critères d'appréciation aussi sûrs, aussi valides que possible. Ces critères ont été appliqués à de nombreux travaux.

Dans l'espoir d'assurer l'homogénéité des notes, des consignes de correction détaillées sont proposées (et justifiées).

L'évaluation à l'aide d'échelles descriptives fournit deux types d'informations à l'élève :

- Il est informé des points forts et des lacunes de son travail;
- Il dispose de points de repère pour situer sa performance par rapport à ses condisciples.

En raison de son caractère analytique, l'évaluation à l'aide d'échelles descriptives prend beaucoup de temps : elle ne pourrait pas être pratiquée quotidiennement dans une classe. Elle devrait être réservée pour certains moments cruciaux de l'apprentissage, ou, plus simplement, pour des vérifications périodiques. Au lieu de se servir des batteries complètes qui lui sont proposées, le professeur a aussi toujours le loisir de n'utiliser qu'une ou deux des échelles, selon les besoins.

On peut espérer que, petit à petit, se créeront des banques d'instruments d'évaluation auxquelles les enseignants pourront faire appel à leur convenance et à celle de leurs élèves.

Notes pour l'évaluation des travaux.

1. Les critères d'évaluation proposés dans les échelles ci-après ont été définis et testés par une commission de professeurs, au cours de nombreuses séances de travail. Il se peut que ces critères ne soient pas ceux en fonction desquels vous jugez habituellement. Pour essayer l'instrument, il importe cependant que vous vous conformiez strictement aux consignes de correction proposées.

En vue de modifications ultérieures seront précieuses.

2. Pour la commodité et l'efficacité de travaux de vos élèves selon la première suivantes.

essus d'enseignement dont elle est une
garantir la *validité* des instruments de
essentatif de la matière dont on souhaite
t à la fois tous les contenus et tous les
ne telle procédure nécessite des études
mentaux de l'acquisition d'une notion.
élève de son progrès vers un objectif
ève peut être comparée à celles d'une
Nous sommes encore loin de disposer

l'évaluation doit être aussi *objective*
ance entre différents juges est un indice
nombreuses études docimologiques ont
nes, combien les juges s'accordent peu
nnée. De plus, l'élève ne peut tirer un
s qui l'accompagnent sont analytiques,
uisse apporter un tel *feedback* à l'élève,
note doit, elle aussi, être analytique.
plus grandes que la performance solli-
nnaissances et d'aptitudes nombreuses.
construites pour l'évaluation de per-
nis leur expérience en commun pour
aussi valides que possible. Ces critères

notes, des consignes de correction

fournit deux types d'informations à

es de son travail;
er sa performance par rapport à ses

luation à l'aide d'échelles descriptives
s être pratiquée quotidiennement dans
s moments cruciaux de l'apprentissage,
iodiques. Au lieu de se servir des bat-
sseur a aussi toujours le loisir de n'uti-
ins.

nt des banque s d'instruments d'évalua-
pel à leur convenance et à celle de leurs

chelles ci-après ont été définis et testés
s de nombreuses séances de travail. Il
n fonction desquels vous jugez habi-
orte cependant que vous vous confor-
a proposées.

En vue de modifications ultérieures des critères, vos remarques et vos critiques seront précieuses.

2. Pour la commodité et l'efficacité de l'évaluation, il est conseillé d'évaluer tous les travaux de vos élèves selon la première échelle, puis de procéder de même pour les suivantes.

A. — Description de l'exercice

a) OBJECTIFS DE L'EXPÉRIENCE.

En début d'année scolaire, des élèves de 4^e année de l'enseignement secondaire rénové ont été invités à :

- observer une expérience;
- interpréter les phénomènes observés;
- imaginer une méthode d'analyse.

Il s'agit d'une expérience de simple substitution : déplacements d'halogènes par d'autres halogènes. Elle est préparatoire à la notion de gradation dans le pouvoir oxydant des éléments et dans les réactions redox de simple substitution.

L'exercice proposé prépare l'acquisition de notions théoriques importantes qui seront développées ultérieurement. On souhaite ici partir d'une situation n'exigeant aucune connaissance préalable du sujet.

b) PRÉSENTATION SCHÉMATIQUE DE L'EXPÉRIENCE.

1. Présentation des produits.

L'expérience proposée aux élèves mettra en présence les produits suivants :

- 3 réactifs corps simples : Iode (I_2)
Brome (Br_2)
Chlore (Cl_2).
- 3 réactifs corps composés : Iodure de potassium (KI)
Bromure de potassium (KBr)
Chlorure de potassium (KCl).
- 2 solvants : Eau (H_2O)
Tétrachlorure de carbone (CCl_4).

2. Manipulations préliminaires.

Pour que cette expérience soit intelligible pour l'élève, il convient de procéder à des manipulations préliminaires. Les résultats de ces manipulations constituent pour l'élève des références indispensables.

- Mise en présence des deux solvants (H_2O et CCl_4);
- Mise en présence des deux solvants avec chacun des réactifs — corps simples;
- Mise en présence des deux solvants avec chacun des réactifs — corps composés.

3. Expérience.

Les réactifs sont pris deux à deux (un corps simple et un corps composé) :

- en solution aqueuse d'abord;
- avec addition de CCl_4 ensuite.

c) CONDITIONS ET DÉROULEMENT DE L'EXPÉRIENCE.

1. Notions prérequisées.

Ce sont les notions que l'élève de

- Réactif - réaction ou phénomène chimique;
- Solvant - solution aqueuse - insoluble;
- Corps simple - Corps composé.
- Élément.

2. Manipulations préliminaires.

Les vingt éprouvettes seront préparées et présentées à l'élève. Le professeur remplira les vingt éprouvettes et complètera au fur et à mesure le tableau ci-après (tableau I).

Les gammes de couleurs sont présentées dans le vocabulaire employé. Il est recommandé de renseigner l'élève sur le contenu. Par

Solvants	H_2O
Réactifs	incolore
Corps simples Br_2	jaune-orangé (3)
I_2	jaune-brun (6)
Cl_2	incolore-jaunâtre (9)
Corps composés KBr	incolore (limpide) (12)
KCl	incolore (limpide) (15)
KI	incolore (limpide) (18)

N.B. — On trouvera, dans les directives, les numéros de certains de ces produits. Les numéros

titution : déplacements d'halogènes par la notion de gradation dans le pouvoir redox de simple substitution.

de notions théoriques importantes qui aite ici partir d'une situation n'exigeant

CE.

en présence les produits suivants :

- potassium (KI)
- le potassium (KBr)
- le potassium (KCl).

ure de carbone (CCl₄).

pour l'élève, il convient de procéder ultats de ces manipulations constituent

- et CCl₄);
- chacun des réactifs — corps simples;
- chacun des réactifs — corps composés.

ps simple et un corps composé) :

c) CONDITIONS ET DÉROULEMENT DE L'EXPÉRIENCE.

1. Notions prérequis.

Ce sont les notions que l'élève devrait maîtriser avant d'aborder cet exercice.

- Réactif - réaction ou phénomène chimique.
- Solvant - solution aqueuse - insolubilité - miscibilité de deux liquides.
- Corps simple - Corps composé.
- Élément.

2. Manipulations préliminaires.

Les vingt éprouvettes seront préparées à l'avance par le professeur (cf. directives) et présentées à l'élève. Le professeur fera un commentaire à propos de chacune des éprouvettes et complètera au fur et à mesure un tableau semblable à celui que l'on trouve ci-après (tableau I).

Les gammes de couleurs sont présentées de manière à supprimer toute ambiguïté dans le vocabulaire employé. Il est recommandé de placer sur chaque tube une étiquette renseignant l'élève sur le contenu. Par exemple, pour l'éprouvette 8 : I₂, H₂O, CCl₄.

TABLEAU I

Solvants	H ₂ O	CCl ₄	H ₂ O + CCl ₄	
			phase sup.	phase inf.
Réactifs	incoloré	incoloré (1)	incoloré (2)	
<i>Corps simples</i> BR ₂	jaune-orangé (3)	rouge (4)	jaune orangé	rouge (5)
I ₂	jaune-brun (6)	violet (7)	jaune brun	violet (8)
Cl ₂	incoloré-jaunâtre (9)	incoloré (10)	incoloré jaunâtre	incoloré jaunâtre(11)
<i>Corps composés</i> KBr	incoloré (limpide) (12)	incoloré+cristaux (13)	incoloré	incoloré(14)
KCl	incoloré (limpide) (15)	incoloré+cristaux (16)	incoloré	incoloré(17)
KI	incoloré (limpide) (18)	incoloré+cristaux (19)	incoloré	incoloré(20)

N.B. — On trouvera, dans les directives ci-après, des notes relatives à la préparation de certains de ces produits. Les numéros entre parenthèses renvoient à ces directives.

REMARQUES :

1. A ce stade du travail, il pourrait être avantageux de faire remarquer aux élèves :
 - que tous les phénomènes rencontrés jusqu'ici sont des phénomènes physiques; les substances rencontrées sont restées intactes;
 - que des réactions peuvent cependant se produire dans les manipulations à venir.
2. Il va de soi que, pour des travaux pratiques échelonnés sur plusieurs séances, la participation directe des élèves serait sollicitée pour les manipulations qui précèdent, et surtout pour celles qui vont suivre.

Directives relatives à la préparation des produits :

Il est conseillé d'utiliser des produits dits « pour analyse ».

- (2) Présenter un tube à essai contenant un volume d'eau et un second, avec un volume moitié moindre de CCl_4 .

Ajouter lentement le CCl_4 à l'eau pour que les élèves le voient descendre au fond du tube à essai et conclure à la non miscibilité des deux solvants.

Agiter et faire constater que le CCl_4 redescend de nouveau ⁽¹⁾.

Si des élèves observent la formation d'un ménisque de CCl_4 au-dessus de la phase aqueuse, leur signaler qu'il s'agit d'un phénomène physique sans intérêt pour l'expérience.

Toutes les solutions décrites ci-après ne doivent pas être préparées trop longtemps d'avance. Leurs concentrations, vérifiées lors d'essais préalables, doivent être telles que les teintes obtenues lors des réactions soient nettes et identifiables sans aucun doute possible.

- (3 - 6 - 9)

Le professeur peut montrer, ou non, l'iode, le brome et le chlore à l'état de corps purs, pour satisfaire la curiosité scientifique des élèves. Comme les manipulations se feront de toute manière avec leurs solutions aqueuses, pour des raisons de sécurité, on peut toutefois se limiter à ne présenter et à ne décrire que ces dernières.

- (3) Utiliser une solution saturée de brome.
- (6) Ajouter quelques cristaux d'iodure de potassium.
- (9) Utiliser une solution saturée de chlore.

- (4 - 5 - 7 - 8 - 10 - 11)

Partir d'une solution aqueuse saturée et ajouter un volume égal de CCl_4 .

Prélever la moitié du CCl_4 (se servir d'une ampoule de décantation) pour préparer les solutions 4 - 7 - 10, et conserver le reste qui constitue les solutions 5, 8 et 11.

- (6) Ajouter quelques cristaux d'iodure de potassium, de manière à augmenter la solubilité de l'iode.
- (9) Utiliser une solution saturée de chlore.

⁽¹⁾ Cette remarque vaut pour toutes les manipulations où intervient le CCl_4 .

(12 - 15 - 18)

Utiliser des solutions à 5 %.

(13 - 16 - 19)

Quelques cristaux dans une même cuvette.

3. *Expérience.*

Les questions posées aux élèves portées sur la page 195. Neuf expériences seront présentées par groupe.

Chacune des expériences sera menée :

- a) en solution aqueuse;
- b) en ajoutant CCl_4 au mélange des deux solvants simple présent à l'issue de la réaction.

AVANT chaque expérience, le professeur doit mettre en œuvre. EXEMPLE : $A_1 = \text{Sol KI} + \text{eau}$

APRÈS chaque expérience, l'élève doit noter les temps de l'expérience.

Groupe A

- $A_1 = \text{Sol KBr} + \text{eau}$
- $A_2 = \text{Sol KBr} + \text{eau}$
- $A_3 = \text{Sol KBr} + \text{eau}$

Groupe B

- $B_1 = \text{Sol KI} + \text{eau}$
- $B_2 = \text{Sol KI} + \text{eau}$
- $B_3 = \text{Sol KI} + \text{eau}$

Groupe C

- $C_1 = \text{Sol KCl} + \text{eau}$
- $C_2 = \text{Sol KCl} + \text{eau}$
- $C_3 = \text{Sol KCl} + \text{eau}$

Les quantités de chacune des deux solutions de carbone doivent être sensiblement égales. Le dernier représente approximativement 10 %.

Le professeur prendra la précaution de faire l'addition de CCl_4 .

Dans les conditions décrites, l'expérience sera menée de la manière suivante :

(12 - 15 - 18)

Utiliser des solutions à 5 %.

(13 - 16 - 19)

Quelques cristaux dans une même quantité de CCl_4 .

3. Expérience.

Les questions posées aux élèves porteront uniquement sur cette partie de l'exercice. Neuf expériences seront présentées par groupe de trois aux élèves.

Chacune des expériences sera menée en deux temps :

- en solution aqueuse;
- en ajoutant CCl_4 au mélange des deux solutions aqueuses pour identifier le corps simple présent à l'issue de la réaction.

AVANT chaque expérience, le professeur indique au tableau les produits qu'il va mettre en œuvre. EXEMPLE : $A_1 = \text{Sol KBr} + \text{eau de chlore}$.

APRÈS chaque expérience, l'élève note ses observations, relativement aux deux temps de l'expérience.

Groupe A

- $A_1 = \text{Sol KBr} + \text{eau de chlore}$
- $A_2 = \text{Sol KBr} + \text{eau de brome}$
- $A_3 = \text{Sol KBr} + \text{eau d'iode (solution d'iodo-iodurée)}$

Groupe B

- $B_1 = \text{Sol KI} + \text{eau de chlore}$
- $B_2 = \text{Sol KI} + \text{eau de brome}$
- $B_3 = \text{Sol KI} + \text{eau d'iode (solution iodo-iodurée)}$

Groupe C

- $C_1 = \text{Sol KCl} + \text{eau de chlore}$
- $C_2 = \text{Sol KCl} + \text{eau de brome}$
- $C_3 = \text{Sol KCl} + \text{eau d'iode (solution iodo-iodurée)}$

Les quantités de chacune des deux solutions de réactifs et celle de tétrachlorure de carbone doivent être sensiblement égales, de sorte que finalement le volume de ce dernier représente approximativement le tiers du total.

Le professeur prendra la précaution d'agiter les tubes à essais spécialement après addition de CCl_4 .

Dans les conditions décrites, l'expérience présentée doit mener aux résultats suivants :

TABLEAU II

Réactifs corps composés	KBr		KI		KCl	
	H ₂ O	CCl ₄	H ₂ O	CCl ₄	H ₂ O	CCl ₄
I ₂	jaune-brun	jaune-brun (1) (2)	jaune-brun	jaune-brun	jaune-brun	jaune-brun
		violet I ₂		violet I ₂		violet I ₂
Br ₂	jaune-orangé	jaune-orangé	jaune-brun	jaune-brun	jaune-orangé	jaune-orangé
		rouge Br ₂		violet I ₂		rouge Br ₂
Cl ₂	jaune-orangé	jaune-orangé	jaune-brun	jaune-brun	incolore	incolore
		rouge Br ₂		violet I ₂		incolore Cl ₂

Ce tableau est présenté à titre indicatif, pour donner une image complète de l'expérience. Il n'est évidemment pas communiqué aux élèves. La présentation des observations en tableau à double entrée n'a d'ailleurs pas été induite par le professeur.

(¹) La disposition adoptée dans le tableau représente les deux phases observées dans l'éprouvette : en haut, la phase aqueuse; en bas, la phase organique.

(²) De légères variations de teinte peuvent s'observer entre la phase aqueuse seule et la phase supérieure après adjonction de CCl₄ : elles s'expliquent par une moindre concentration des réactifs dans la phase supérieure en CCl₄.

4. Questions posées aux élèves.

1. Qu'observez-vous dans l'eau et dans le CCl₄ ?
2. Pour chaque expérience, expliquer les modifications de couleur observées.
3. Ecrivez les équations rendant compte des phénomènes chimiques observés.
4. a) Quel est l'atome le moins bien « accroché » au potassium ? Justifiez.
b) Quel est l'atome le mieux « accroché » au potassium ? Justifiez.
c) Etablissez un classement des trois éléments suivant leur force décroissante.
5. Une solution contient peut-être de l'iode; imaginez une expérience qui permettrait de le vérifier.

B. — Justification

L'échelle d'évaluation est un instrument complexe pour lesquelles un simple décompte

Parmi les performances complexes du secondaire, relater et interpréter une expérience d'apprentissage fondamental.

C. — Choix de

- 1° Relation des phénomènes observés.
- 2° Interprétation des phénomènes observés.
- 3° Invention d'une méthode d'analyse.

4. Questions posées aux élèves.

KCl		
CCl ₄	H ₂ O	CCl ₄
jaune-brun	jaune-brun	jaune-brun
violet I ₂		violet I ₂
jaune-brun	jaune-orangé	jaune-orangé
violet I ₂		rouge Br ₂
jaune-brun	incolore	incolore
violet I ₂		incolore Cl ₂

pour donner une image complète de l'exercice aux élèves. La présentation des observations pas été induite par le professeur.

1. Qu'observez-vous dans l'eau et dans le CCl ₄ ?	Le professeur laisse à l'élève le choix de la présentation de sa réponse.
2. Pour chaque expérience, expliquer les modifications de couleur observées.	Le professeur précisera qu'il ne s'agit pas ici d'écrire les équations.
3. Ecrivez les équations rendant compte des phénomènes chimiques observés.	Le professeur écrira au tableau, sous la forme du premier membre d'une équation, les réactifs mis en évidence. A1 : KBr + Cl ₂ A2 : KBr + Br ₂ A3 : KBr + I ₂ B1 : KI + Cl ₂ B2 : KI + Br ₂ B3 : KI + I ₂ C1 : KCl + Cl ₂ C2 : KCl + Br ₂ C3 : KCl + I ₂
4. a) Quel est l'atome le moins bien « accroché » au potassium ? Justifiez. b) Quel est l'atome le mieux « accroché » au potassium ? Justifiez. c) Etablissez un classement des trois éléments suivant leur force décroissante.	Le terme approximatif « accroché » est intentionnel : l'exercice proposé n'est qu'une approche de notions théoriques importantes qui seront développées ultérieurement.
5. Une solution contient peut-être de l'iodure; imaginez une expérience qui permettrait de le vérifier.	

B. — Justification du choix de cet exercice

L'échelle d'évaluation est un instrument destiné à évaluer des performances complexes pour lesquelles un simple décompte de points est inadéquat.

Parmi les performances complexes que l'on peut envisager dans l'enseignement secondaire, relater et interpréter une expérience est, de l'avis des professeurs, un apprentissage fondamental.

C. — Choix des critères d'évaluation

- 1° Relation des phénomènes observés.
- 2° Interprétation des phénomènes observés.
- 3° Invention d'une méthode d'analyse.

représente les deux phases observées dans le bas, la phase organique. Elles s'observent entre la phase aqueuse seule et la phase CCl₄ : elles s'expliquent par une moindre solubilité de l'iodure en CCl₄.

L'évaluation selon ces trois critères n'est évidemment pas exhaustive. Elle porte sur ce qui est spécifique à ce genre d'exercice et non sur sa présentation formelle (style, orthographe, etc.).

D. — Organisation du dossier

Pour chacun des critères retenus, on présentera successivement :

- La définition;
- La définition des aspects intervenant;
- La détermination des niveaux d'exigence.

Les exigences correspondant à chaque degré d'une échelle ne sont pas arbitraires; elles ont été établies après une analyse approfondie de nombreux travaux de quatrième année de l'enseignement secondaire rénové.

- La démarche d'évaluation;
- L'échelle.

Pour évaluer, l'échelle n'apporte aucune information complémentaire par rapport à la démarche déjà décrite. Toutefois, l'échelle offre l'avantage de donner une description hiérarchique de la performance.

PRÉSENTATION

I. — RELATION

A. — Définition

Relater une expérience, c'est traduire par un plan, croquis, tableau à double entrée...

B. — Définition des aspects

On évaluera deux aspects :

- La relation des phénomènes essentiels et les informations minimales, suffisantes, évaluées.

- Pour les situations correctement relatées, la relation. On valorise ici des éléments de l'interprétation de l'expérience, et non pas l'expérience elle-même.

Ces deux aspects auront le même poids dans l'évaluation de l'expérience.

1. Relation des phénomènes essentiels

Un travail complet comportera la description de la phase la plus haute.

Pour l'évaluation, on ne distinguera pas les phases.

On considérera qu'une expérience est correctement relatée si l'élève décrit les phénomènes observés.

- en solution aqueuse (soit en solution aqueuse, soit après adjonction de CCl_4 (*) ;
- en CCl_4 (phase inférieure).

Pour évaluer chacune des expériences présentées dans le tableau II.

2. Précision.

L'évaluation de la précision ne portera que sur la phase la plus haute.

a) Distinction entre phase aqueuse et phase inférieure

Un travail précis établit une distinction entre la phase aqueuse et la phase inférieure, après adjonction de CCl_4 . En l'absence de CCl_4 , la phase inférieure est la phase aqueuse.

(*) Si, pour une expérience, le travail présente des couleurs en solution aqueuse, et qu'il n'y a pas d'erreur, pour autant qu'il n'y a pas de discordance sera pénalisée par ailleurs.

PRÉSENTATION DES ÉCHELLES

I. — RELATION DE L'EXPÉRIENCE

A. — Définition du critère

Relater une expérience, c'est traduire par un moyen quelconque (texte continu, plan, croquis, tableau à double entrée, etc.) des phénomènes observés.

B. — Définition des aspects qui interviennent dans le critère

On évaluera deux aspects :

— La relation des phénomènes essentiels à l'interprétation de l'expérience. Seules les informations minimales, suffisant à la compréhension de l'expérience sont évaluées.

— Pour les situations correctement relatées, on évaluera en outre la *précision de cette relation*. On valorise ici des éléments du travail de l'élève qui, sans être essentiels à l'interprétation de l'expérience, traduisent une observation plus rigoureuse.

Ces deux aspects auront le même poids dans l'évaluation de la relation de l'expérience.

1. Relation des phénomènes essentiels à l'interprétation de l'expérience.

Un travail complet comportera la description correcte des neuf expériences décrites plus haut.

Pour l'évaluation, on ne distingue pas *erreur et omission*.

On considérera qu'une expérience est complète et correctement relatée si le travail de l'élève décrit les phénomènes observés à la fois :

— en solution aqueuse (soit en solution aqueuse seule, soit dans la phase supérieure après adjonction de CCl_4) ⁽¹⁾;

— en CCl_4 (phase inférieure).

Pour évaluer chacune des expériences, on se reportera aux *gammes de couleurs* présentées dans le tableau II.

2. Précision.

L'évaluation de la précision ne porte que sur les situations correctement décrites.

a) Distinction entre phase aqueuse et phase organique.

Un travail précis établit une distinction entre la phase inférieure et la phase supérieure, après adjonction de CCl_4 . En outre, il faut que l'élève localise au moins une fois le CCl_4 dans la phase inférieure.

⁽¹⁾ Si, pour une expérience, le travail présente une discordance dans l'identification des couleurs en solution aqueuse, et en phase supérieure en CCl_4 , on considérera qu'il n'y a pas d'erreur, pour autant qu'une des deux observations soit exacte. Cette discordance sera pénalisée par ailleurs : absence de répétition (voir page 200).

b) Répétition des phénomènes observés en solution aqueuse.

On valorisera tout travail qui, après avoir décrit correctement les phénomènes observés en solution aqueuse seule, mentionne la phase supérieure après adjonction du CCl_4 .

La répétition n'est possible que si l'élève établit préalablement la distinction entre phase aqueuse et phase organique.

C. — Détermination des niveaux d'exigence

a) Pour le premier aspect « Relation des phénomènes essentiels », on retiendra cinq niveaux :

- Aucune erreur ou omission
- Une erreur ou omission
- Deux erreurs ou omissions
- Plus de deux erreurs ou omissions
- Aucune référence à l'une des deux phases.

b) Pour le deuxième aspect « Précision », on retiendra les niveaux suivants :

- Distinction : Oui
Non
- Répétition : Oui
Non

D. — Démarche pour l'évaluation du critère : « Relation des phénomènes observés »

1) Appréciez le premier aspect : *Relation des phénomènes essentiels*.

- Aucune erreur
- Une erreur ⁽¹⁾
- Deux erreurs ⁽¹⁾
- Plus de deux erreurs ⁽¹⁾
- Aucune référence à l'une des deux phases

2) Appréciez le deuxième aspect : *Précision*.

- a) Distinction
- Pas de distinction
- b) Répétition
- Pas de répétition

⁽¹⁾ Indiquez le(s) code(s) des situations omises ou incorrectes.

3. En vous reportant au schéma suivant, indiquez les caractéristiques que vous venez d'observer :

Relation des phénomènes essentiels	Distinction Localisation	Précision
— Aucune erreur	{ — Oui } { — Non }	{ — Oui } { — Non }
— Une erreur	{ — Oui } { — Non }	{ — Oui } { — Non }
— Deux erreurs	{ — Oui } { — Non }	{ — Oui } { — Non }
— Plus de deux erreurs		
— Aucune référence à l'une des deux phases		

E. — Présentation de l'évaluation

	Relation des phénomènes observés
TB	Aucune erreur
B	Aucune erreur
	Une erreur
S	Aucune erreur
	Une erreur
	Deux erreurs
F	Une erreur
	Deux erreurs
I	Deux erreurs
	Plus de deux erreurs Aucune référence à l'une des deux phases

lution aqueuse.
 décrit correctement les phénomènes
 la phase supérieure après adjonction

blit préalablement la distinction entre

veaux d'exigence

mènes essentiels », on retiendra cinq

endra les niveaux suivants :

Relation des phénomènes observés »

phénomènes essentiels.

-
-
-
-
-
-
-
-
-
-

s ou incorrectes.

3. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux trois caractéristiques que vous venez d'évaluer dans le travail de l'élève.

Relation des phénomènes essentiels	Précision		
	Distinction Localisation	Répétition	
— Aucune erreur	— Oui	— Oui	TB
	— Non	— Non	B
— Une erreur	— Oui	— Non	S
	— Non	— Non	B
— Deux erreurs	— Oui	— Non	S
	— Non	— Non	F
— Plus de deux erreurs			I
— Aucune référence à l'une des deux phases			F
			I

Note pour la Relation des phénomènes essentiels :

E. — Présentation de l'échelle : « Relation de l'expérience »

	Relation des phénomènes essentiels	Précision	
		Distinction Localisation	Répétition
TB	Aucune erreur	Oui	Oui
B	Aucune erreur	Oui	Non
	Une erreur	Oui	Oui
S	Aucune erreur	Non	Non
	Une erreur	Oui	Non
	Deux erreurs	Oui	Oui
F	Une erreur	Non	Non
	Deux erreurs	Oui	Non
I	Deux erreurs	Non	Non
	Plus de deux erreurs Aucune référence à l'une des deux phases	Quelles que soient les caractéristiques de précision	

II. — INTERPRÉTATION DES PHÉNOMÈNES OBSERVÉS

A. — Définition du critère

Interpréter, c'est mettre en relation les phénomènes observés pour leur donner une signification.

B. — Définition des aspects qui interviennent dans le critère

Les réponses aux questions 2, 3 et 4 participent d'un même processus; elles seront donc évaluées sur une même échelle.

L'évaluation portera sur trois aspects :

- L'identification des corps simples présents dans le milieu à l'issue des expériences;
- La distinction entre phénomènes physiques et réactions chimiques;
- Les conclusions sur la force relative des halogènes.

1° IDENTIFICATION DES CORPS SIMPLES.

L'identification correcte des corps simples, présents dans le milieu à l'issue des expériences, suppose que l'élève établisse une relation correcte entre la couleur observée dans le CCl_4 et la nature du corps simple.

- Pour trois expériences ($\text{KI} + \text{Br}_2$ — $\text{KI} + \text{Cl}_2$ — $\text{KBr} + \text{Cl}_2$), le corps à identifier s'est formé à la suite d'une réaction chimique (expérience de type I).
- Pour les six autres expériences, le corps à identifier était déjà introduit dans la manipulation : il n'y a pas eu de réaction chimique (expérience de type II).

Ces deux catégories d'expériences se verront accorder une importance différente dans l'échelle.

2° DISTINCTION ENTRE PHÉNOMÈNES PHYSIQUES ET RÉACTIONS CHIMIQUES.

L'élève peut établir cette distinction de deux manières : par un texte écrit ou par des équations :

a) Par écrit.

- Pour chaque expérience, l'élève fait allusion à la présence ou à l'absence du corps identifié dans les manipulations et conclut qu'il y a ou non réaction. EXEMPLE :

	<i>Identification</i>	<i>Distinction</i>
. $\text{KCl} + \text{Br}_2$	rouge ↓ Brome	Br_2 introduit dans la manipulation → pas de réaction
. $\text{KI} + \text{Br}_2$	violet ↓ Iode	I_2 n'a pas été introduit dans la manipulation → réaction

ou

202

- Pour chaque expérience, l'élève co

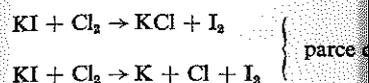
	<i>Identification</i>
. $\text{KCl} + \text{Br}_2$	rouge ↓ Brome
. $\text{KI} + \text{Br}_2$	violet ↓ Iode

b) Par les équations.

L'élève fait la preuve qu'il établit, chimiques et phénomènes physiques s

- il ne complète pas le deuxième me
- riences de type II,
- et fait apparaître le corps formé

On n'évalue pas ici la correction donc comme équivalentes et comme p



Par contre, on rejette $\text{KI} + \text{Cl}_2 \rightarrow \text{I}$ formé.

3° CONCLUSIONS SUR LA FORCE RELA

Cet aspect est relatif à l'évaluation

a) Classement.

On considérera que l'élève établit réponse aux questions 4a, 4b et 4c négativement.

b) Justification.

Pour être correctes, les justification I_2 est le moins bien accroché, car l'
ou

Cl_2 est le mieux accroché, car le chl

L'incorrection de l'une des deux p travail. Par contre, les deux justificati elles est tolérée.

PHÉNOMÈNES OBSERVÉS

du critère

phénomènes observés pour leur donner une

interviennent dans le critère

peuvent d'un même processus; elles seront

dans le milieu à l'issue des expériences;

réactions chimiques;

halogènes.

présents dans le milieu à l'issue des

relation correcte entre la couleur observée

 $\text{Cl}_2 - \text{KBr} + \text{Cl}_2$, le corps à identifier
 chimique (expérience de type I).

 à identifier était déjà introduit dans la
 chimique (expérience de type II).

on peut accorder une importance différente

PHÉNOMÈNES ET RÉACTIONS CHIMIQUES.

manières : par un texte écrit ou par des

 on se base sur la présence ou à l'absence du corps
 chimique qu'il y a ou non réaction. EXEMPLE :
Distinction
 $\text{KCl} + \text{Br}_2$ introduit dans la manipulation

→ pas de réaction

 $\text{KI} + \text{Br}_2$ n'a pas été introduit dans la manipu-
 lation

→ réaction

Pour chaque expérience, l'élève conclut à une réaction ou non. EXEMPLE :

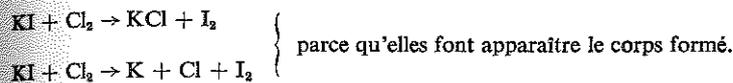
	Identification	Distinction
$\text{KCl} + \text{Br}_2$	rouge ↓ Brome	Pas de réaction
$\text{KI} + \text{Br}_2$	violet ↓ Iode	Réaction

b) Par les équations.

 L'élève fait la preuve qu'il établit, par les équations, la distinction entre réactions
 chimiques et phénomènes physiques si, parmi les 9 équations dont on lui a proposé le
 premier membre :

 — il ne complète pas le deuxième membre ou répète le premier pour toutes les expé-
 riences de type II,

— et fait apparaître le corps formé pour les trois expériences de type I.

 On n'évalue pas ici la correction et l'équilibrage des équations; on considère
 donc comme équivalentes et comme preuve de la distinction les équations suivantes :

 Par contre, on rejette $\text{KI} + \text{Cl}_2 \rightarrow \text{KICl}_2$ parce qu'elle ne fait pas apparaître le corps
 formé.

3° CONCLUSIONS SUR LA FORCE RELATIVE DES HALOGÈNES.

Cet aspect est relatif à l'évaluation des questions 4a, 4b, 4c.

a) Classement.

 On considérera que l'élève établit correctement le classement des halogènes si la
 réponse aux questions 4a, 4b et 4c est correcte. Tous les autres cas seront évalués
 négativement.

b) Justification.

Pour être correctes, les justifications devront comporter les informations suivantes :

 I_2 est le moins bien accroché, car l'iode est déplacé par Br_2 et Cl_2

ou

 Cl_2 est le mieux accroché, car le chlore n'est pas déplacé par Br_2 .

 L'incorrection de l'une des deux justifications entraîne une évaluation négative du
 travail. Par contre, les deux justifications étant équivalentes, l'omission de l'une d'entre
 elles est tolérée.

C. — Détermination des niveaux d'exigence

- a) Pour l'aspect « Identification des corps simples », on retiendra trois niveaux :
- Les corps sont correctement identifiés pour les trois expériences du type I et pour cinq expériences du type II au moins (on tolère donc une erreur pour ce type d'expérience);
 - Les corps sont correctement identifiés pour les trois expériences du type I et pour moins de cinq expériences du type II;
 - Les corps pour les trois expériences du type I ne sont pas tous correctement identifiés.
- b) Pour l'aspect « Distinction entre phénomènes physiques et réactions chimiques », on retiendra deux niveaux :
- Oui
 - Non.
- c) Pour l'aspect « Conclusions sur la force relative des halogènes », on retiendra les niveaux suivants :
- Classement correct
 - incorrect;
 - Justification correcte
 - incorrecte.

D. — Démarche pour l'évaluation du critère : « Interprétation des phénomènes observés »

1. Appréciez le premier aspect : *Identification des composantes.*

- Les corps sont correctement identifiés :
 - . pour les trois expériences du type A et pour au moins cinq expériences du type B
- Les corps sont correctement identifiés :
 - . pour les trois expériences du type A
 - . pour moins de cinq expériences du type B
- Les corps, pour les trois expériences du type A ne sont pas tous correctement identifiés

2. Appréciez le deuxième aspect : *Distinction entre phénomènes physiques et réactions chimiques* (par le texte ou par les équations).

L'élève établit la distinction :

- Oui
- Non

3. Appréciez le troisième aspect : Co

- a) Classement
- correct
 - incorrect

- b) Justification : l'élève apporte une
- oui
 - non

4. En vous reportant au schéma suivant, indiquez les caractéristiques que vous venez d'é

<i>Identification</i>	<i>Distinction</i>
- Les corps sont correctement identifiés pour les trois expériences du type I et pour au moins cinq expériences du type II	- L'élève a distingué la différence entre les phénomènes physiques et chimiques
- Les corps pour les 3 expériences du type I ne sont pas tous correctement identifiés	- L'élève n'a pas distingué la différence entre les phénomènes physiques et chimiques
- Les corps sont correctement identifiés pour les trois expériences du type I et pour moins de cinq expériences du type II	- L'élève a distingué la différence entre les phénomènes physiques et chimiques
- Les corps pour les 3 expériences du type I ne sont pas tous correctement identifiés	- L'élève n'a pas distingué la différence entre les phénomènes physiques et chimiques

Note pour

niveaux d'exigence

mples », on retiendra trois niveaux :
pour les trois expériences du type I
moins (on tolère donc une erreur pour ce

pour les trois expériences du type I et
II;
type I ne sont pas tous correctement

physiques et réactions chimiques ».

relative des halogènes », on retiendra les

Interprétation des phénomènes observés »

n des composantes.

A et pour au moins cinq expé-

pe B

u type A ne sont pas tous correc-

entre phénomènes physiques et réactions
).

3. Appréciez le troisième aspect : *Conclusions sur la force relative des halogènes.*

a) Classement

- correct
- incorrect

b) Justification : l'élève apporte une justification correcte au moins

- oui
- non

4. En vous reportant au schéma suivant, choisissez le niveau qui correspond aux caractéristiques que vous venez d'évaluer dans le travail de l'élève.

Identification	Force relative des halogènes		
	Distinction	Classement	Justification
- Les corps sont correctement identifiés pour les trois expériences du type I et pour au moins cinq expériences du type II	- L'élève établit la distinction	- correct	- correcte.. TB - incorrecte B
		- incorrect	- correcte.. B - incorrecte S
	- L'élève n'établit pas la distinction	- correct	- correcte.. B - incorrecte S
		- incorrect	- correcte.. S - incorrecte F
- Les corps sont correctement identifiés pour les trois expériences du type I et pour moins de cinq expériences du type II	- L'élève établit la distinction	- correct	- correcte.. S - incorrecte F
		- incorrect	- correcte.. F - incorrecte I
	- L'élève n'établit pas la distinction	- correct	- correcte.. F - incorrecte I
		- incorrect	- correcte.. I - incorrecte I
- Les corps pour les 3 expériences du type I ne sont pas tous correctement identifiés			I

Note pour l'Interprétation des phénomènes observés :

	Identification des composantes	Distinction	Force relative des halogènes	
			Classement	Justification
TB	Les corps sont correctement identifiés pour les 3 expériences du type I et pour au moins 5 expériences du type II	Distinction établie	correct	correcte
B	Les corps sont correctement identifiés pour les 3 expériences du type I et pour au moins 5 expériences du type II	Distinction établie	correct incorrect	incorrecte correcte
		Distinction non établie	correct	correcte
S	Les corps sont correctement identifiés pour les 3 expériences du type I et pour au moins 5 expériences du type II	Distinction établie	incorrect	incorrecte
		Distinction non établie	correct incorrect	incorrecte correcte
F	Les corps sont correctement identifiés pour les 3 expériences du type I et pour au moins 5 expériences du type II	Distinction établie	correct	correcte
	Les corps sont correctement identifiés pour les 3 expériences du type I et pour moins de 5 expériences de type II	Distinction établie	correct incorrect	incorrecte correcte
		Distinction non établie	correct	correcte
I	Les corps sont correctement identifiés pour les 3 expériences du type I et pour moins de 5 expériences du type II	Distinction établie	incorrect	incorrecte
		Distinction non établie	correct	incorrecte
			incorrect incorrect	correcte incorrecte
	Les corps, pour les trois expériences du type I, ne sont pas tous correctement identifiés	Quelles que soient les caractéristiques attribuées pour la distinction et les conclusions		

La démarche d'invention suppose, au-delà des phénomènes, la sélection des variables et la formulation d'un schéma général.

Pour l'exercice proposé, l'expérience doit présenter les caractéristiques suivantes :

1. Mettre en présence la solution à identifier.
2. Mentionner l'addition d'un solvant.
3. Préciser la couleur qui permet de distinguer les corps.

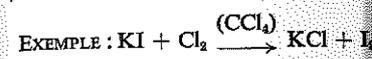
Ces caractéristiques peuvent apparaître dans l'équation.

a) Par le texte.

EXEMPLE : On met en présence du chlorure de potassium et de l'iodure de potassium.

On ajoute CCl_4 . Si la solution devient violette, il y a de l'iodure.

b) Par l'équation.



L'élève ne doit pas nécessairement formuler l'équation, mais qu'il réutilise un des trois corps simples.

B. — Description de l'échelle :

TB	Le travail présenté est correct
B	Le travail présenté est partiellement correct
S	Le travail présenté est incorrect
F	Le travail ne présente aucune caractéristique
I	Le travail ne présente aucune caractéristique

A. — Définition du critère

La démarche d'invention suppose, au-delà de l'observation et de l'interprétation de phénomènes, la sélection des variables pertinentes permettant la mise en place d'un schéma général.

Pour l'exercice proposé, l'expérience imaginée par l'élève devrait présenter les caractéristiques suivantes :

1. Mettre en présence la solution à identifier avec un réactif, soit Br₂, soit Cl₂.
2. Mentionner l'addition d'un solvant (CCl₄).
3. Préciser la couleur qui permet de tirer une conclusion de l'expérience.

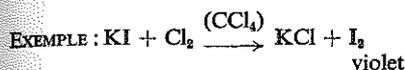
Ces caractéristiques peuvent apparaître sous forme de texte ou sous forme d'équation.

a) Par le texte.

EXEMPLE : On met en présence du chlore et une solution qui contient peut-être un iodure.

On ajoute CCl₄. Si la solution devient violette, c'est qu'elle contenait un iodure.

b) Par l'équation.



L'élève ne doit pas nécessairement faire appel à un nouveau corps. Il est normal qu'il réutilise un des trois corps simples déjà présentés.

B. — Description de l'échelle : « Invention d'une méthode d'analyse »

TB	Le travail présente les caractéristiques 1, 2 et 3
B	Le travail présente les caractéristiques 1 et 2 seulement
S	Le travail présente la caractéristique 1 seulement
F	Le travail ne présente pas la caractéristique 1
I	Le travail ne présente aucune de ces caractéristiques

Note pour le critère : Invention

Distinction	Force relative des halogènes	
	Classement	Justification
Distinction établie	correct	correcte
Distinction établie	correct	incorrecte
	incorrect	correcte
Distinction établie	correct	correcte
Distinction établie	incorrect	incorrecte
Distinction établie	correct	incorrecte
	incorrect	correcte
Distinction établie	correct	correcte
Distinction établie	incorrect	incorrecte
	incorrect	correcte
Distinction établie	correct	incorrecte
	incorrect	incorrecte

elles que soient les caractéristiques attribuées pour la distinction et les conclusions

SYNTHÈSE DES ÉVALUATIONS

Jusqu'ici, vous avez noté les travaux successivement pour la relation et l'interprétation des phénomènes observés, ainsi que pour l'invention d'une méthode d'analyse.

Si l'attribution d'une seule note globale peut détruire le bénéfice d'une démarche analytique préalable, la pratique scolaire impose parfois l'utilisation d'une note de synthèse *complémentaire* à des évaluations analytiques.

Pour cette note de synthèse, le professeur peut selon les objectifs poursuivis accorder la même importance à chacun des critères ou donner plus de poids à telle ou telle appréciation analytique. L'élève devrait être informé de cette décision.

1. Relation					
	TB	B	S	F	I
2. Interprétation					
	TB	B	S	F	I
3. Invention					
	TB	B	S	F	I
Note de synthèse					
	TB	B	S	F	I

essivement pour la relation et l'interprétation de l'invention d'une méthode d'analyse. On ne peut détruire le bénéfice d'une démarche que par la pose parfois l'utilisation d'une note de synthèse analytiques.

On peut selon les objectifs poursuivis accorder plus ou donner plus de poids à telle ou telle information informé de cette décision.

I

I

I

I

ANNEXE B

RÉSULTATS DÉTAILLÉS

HYPOTHÈSE 1 : L'ÉCHELLE ÉVALUATIVE

1. — CONCORDANCE DES NOTES CONSIDÉRÉES
- Tableaux des moyennes et des écarts
- Histogrammes de fréquence.

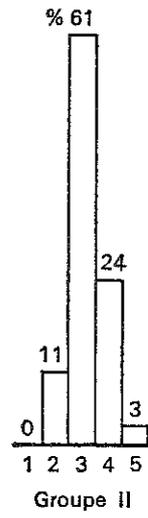
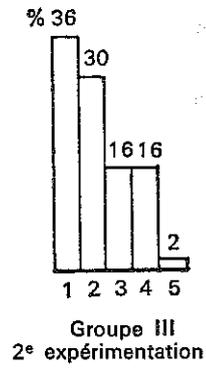
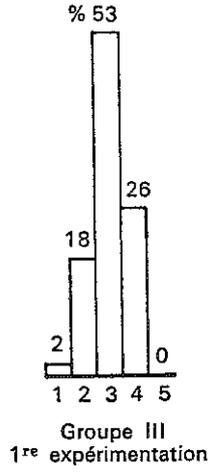
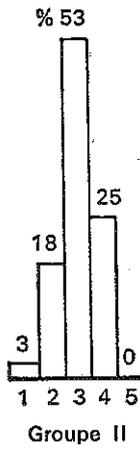
Voir ta

HYPOTHÈSE 1 : L'ÉCHELLE DESCRIPTIVE PERMET UNE
ÉVALUATION PLUS OBJECTIVE

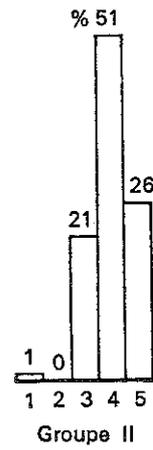
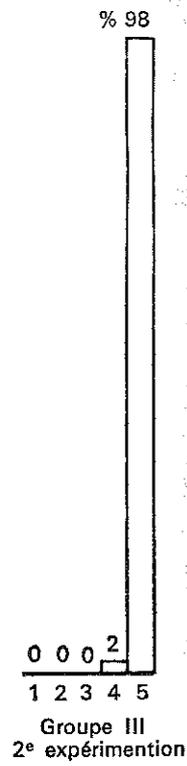
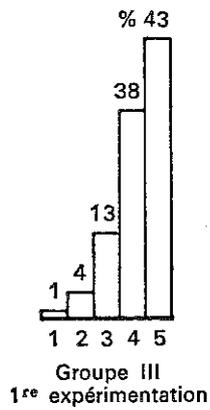
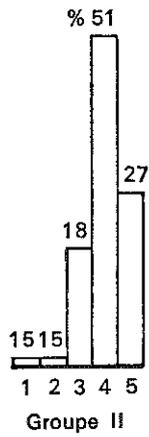
1. — CONCORDANCE DES NOTES CONSIDÉRÉES ISOLÉMENT.
- Tableaux des moyennes et des écarts-types.
- Histogrammes de fréquence.

Voir tableaux ci-après

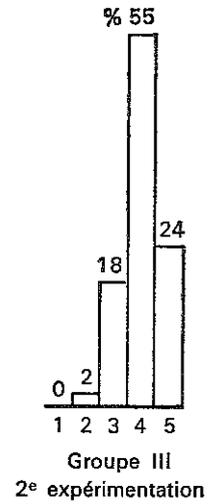
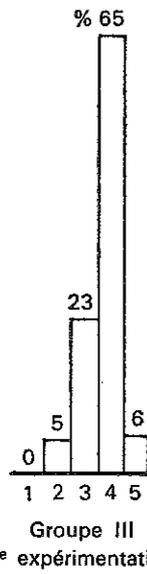
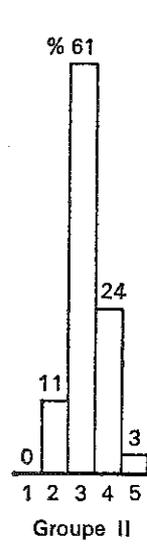
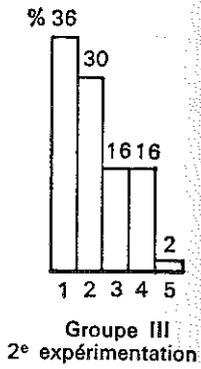
Vocabulaire



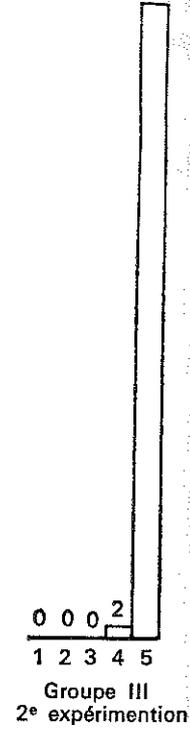
Pertinence syntaxique



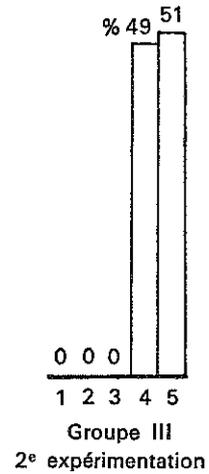
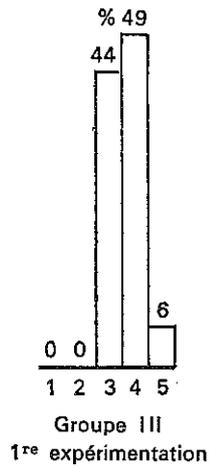
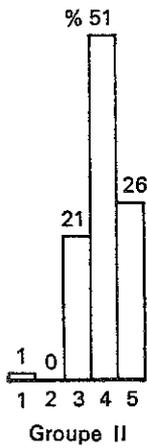
Structures grammaticales



% 98

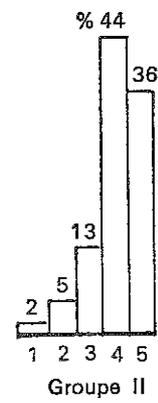
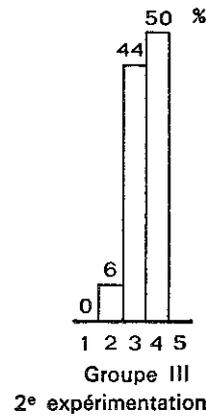
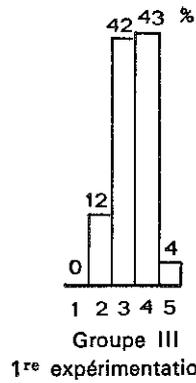
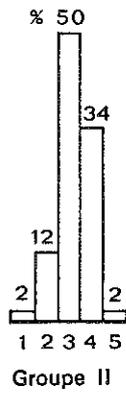
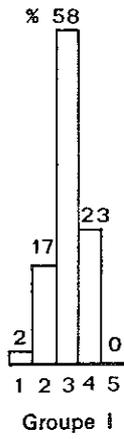


Organisation des idées



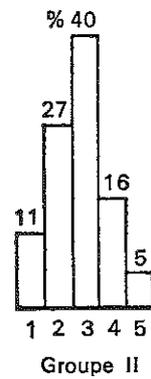
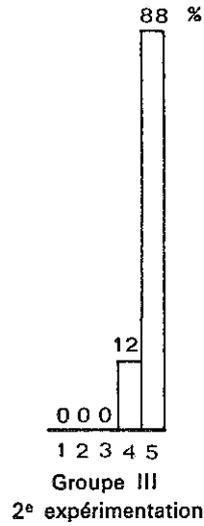
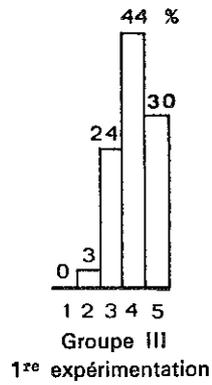
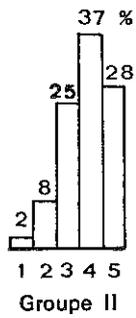
B. SCIENCES

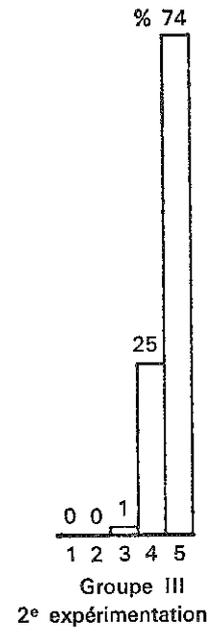
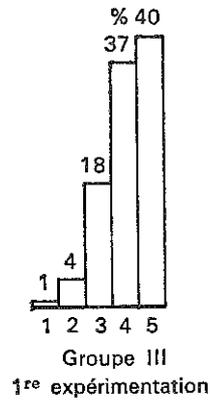
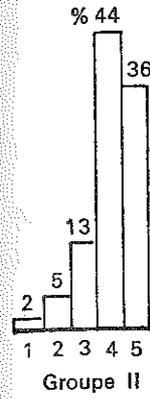
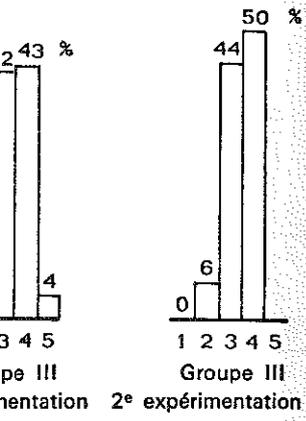
1. Appréciations globales



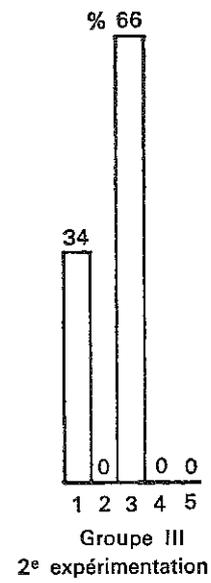
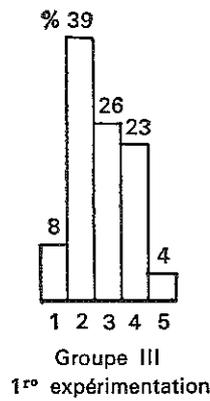
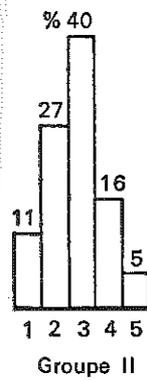
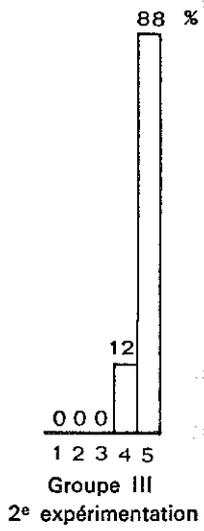
2. Appréciations analytiques

Description

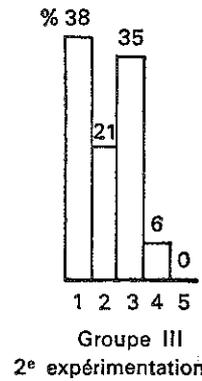
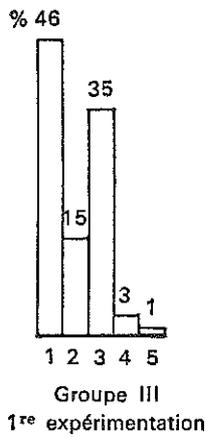
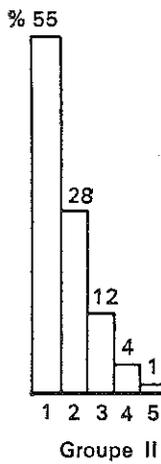




Interprétation



Invention



2. — CONCORDANCE ENTRE LES PROFILS DE NOTES.

a) *Concordance avec le profil de base des notes imposées par les critères de correction.*

	1 ^{re} expérimentation		2 ^e expérim.
	Groupe II	Groupe III	Groupe III
Langue maternelle (5 notes, y compris la note « Vocabulaire »)	1,43 %	1,23 %	2,42 %
Langue maternelle (4 notes, sans la note « Vocabulaire »)	1,43 %	10,11 %	9,23 %
Sciences	1,80 %	1,77 %	1,46 %

Pourcentage d'accord avec le profil de base.

b) *Concordance avec des profils proches du profil de base.*

Les profils retenus présentent, pour un ou deux critères seulement, des valeurs proches (une unité en plus ou en moins) de la note théorique.

Langue maternelle (4 notes) . . .	
Sciences	

3. UNE ÉVALUATION OBJECTIVE N'EST

a) *Les échelles descriptives aident-elles*

1. Evaluation analytique – Analyse de . . .
Voir tableau des moyennes et hist
2. Analyse des notes globales.

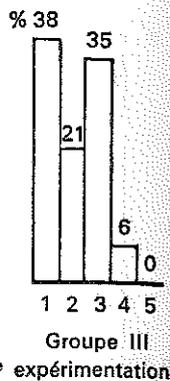
Comparaison globale

Sources de variation	Somme des carrés	D
Inter-groupe (1) . . .	12,019	
Intra-groupe (2) . .	94,506	
Total	106,525	

Langue maternelle

(1) La variance globale inter-groupes d'évaluation utilisées dans les trois . . .
La variance inter-groupe rend com . . .
expérimentaux. Ici, l'introduction de . . .
et des groupes II et III) et l'utilisati . . .
raison des groupes II et III).

(2) La variance intra-groupe cerne . . .



	1 ^{re} expérimentation		2 ^e expérim.
	Groupe II	Groupe III	Groupe III
Langue maternelle (4 notes) ...	34,29	60,50	63,05
Sciences	24,33	34,51	67,60

3. UNE ÉVALUATION OBJECTIVE N'EST PAS INFLUENCÉE PAR DES FACTEURS EXTERNES.

a) Les échelles descriptives aident-elles à lutter contre la tendance centrale ?

1. Evaluation analytique – Analyse des notes attribuées pour chacune des dimensions
Voir tableau des moyennes et histogrammes de fréquence.

2. Analyse des notes globales.

Comparaison globale

Sources de variation	Somme des carrés	Degrés de liberté	Carrés moyens	Valeur de F	Niveaux de signification
Inter-groupe (1) ...	12,019	2	6,01	19,38	S à P.01
Intra-groupe (2) ..	94,506	300	0,31		
Total	106,525	302			

Langue maternelle

1 ^{re} expérimentation	2 ^e expérim.
Groupe III	Groupe III
1,23 %	2,42 %
10,11 %	9,23 %
1,77 %	1,46 %

(1) La variance globale inter-groupe est déterminée par les différentes méthodes d'évaluation utilisées dans les trois groupes.

La variance inter-groupe rend compte des différences éventuelles dues aux effets expérimentaux. Ici, l'introduction de dimensions analytiques (comparaison du groupe I et des groupes II et III) et l'utilisation d'échelles descriptives d'évaluation (comparaison des groupes II et III).

(2) La variance intra-groupe cerne les effets dus au hasard de l'échantillonnage.

Sources de variation	Somme des carrés	Degrés de liberté	Carrés moyens	Valeur de F	Niveaux de signification
Inter-groupe	7,411	2	3,705	7,455	S à P.01
Intra-groupe	180,688	363	0,497		
Total	188,099				

Sciences

Partition des variances

Sources de variation	Somme des carrés	Degrés de liberté	Carrés moyens	Valeurs de F	Niveaux de signification
Inter-groupe (I/II, III)	10,505	1	10,505	33,887	TS à P.01
Inter-groupe (II/III)	1,514	1	1,514	4,883	S à P.01
Intra-groupe	94,506	300	0,31		
Total	106,525				

Langue maternelle

Sources de variation	Somme des carrés	Degrés de liberté	Carrés moyens	Valeurs de F	Niveaux de signification
Inter-groupe (I/II, III)	6,070	1	6,070	12,214	TS
Inter-groupe (II/III)	1,341	1	1,341	2,69	S à P.05
Intra-groupe	180,688	363	0,497		
Total	188,099				

Sciences

3. — ANALYSE DES PROFILS DE NOTES.

a) Profils de tendance centrale.

Nous relevons d'abord le pourcentage qui comporte au moins trois notes centrales.

Langue maternelle

b) Distribution des notes extrêmes.

Langue maternelle

AG	PS	Ps	SC
4	3	5	3
3	4		4
2			même prof
5			même prof
4	4	4	3
3		5	4
2			
5			
4	4	5	3
3			4
2			
5			

Pourcentage d'apparition des profils de notes extrêmes en langue maternelle.

(¹) En sciences, la fréquence d'apparition de ces profils est quasi nulle dans les deux groupes. Les contrastes accusés existant dans le

rrés yens	Valeur de F	Niveaux de signification
705	7,455	S à P.01
497		

rrés yens	Valeurs de F	Niveaux de signification
505 514	33,887 4,883	TS à P.01 S à P.01
31		

rrés yens	Valeurs de F	Niveaux de signification
070 341	12,214 2,69	TS S à P.05
497		

3. — ANALYSE DES PROFILS DE NOTES.

a) Profils de tendance centrale.

Nous relevons d'abord le pourcentage de professeurs dont le profil des évaluations comporte au moins *trois notes centrales* sur les quatre notes attribuées (1).

	1 ^{re} expérimentation		2 ^e expérimentation
	II	III	III
Langue maternelle	10 %	9,87 %	2,1 %

b) Distribution des notes extrêmes.

Langue maternelle

AG	PS	Ps	SG	Org.	Gr. II %	Gr. III %
4	3	5	3	3	2,86	14,80
3	4		4	4	2,86	13,58
2		même profil			0	0
5		même profil			0	2,47
					5,72	30,86
4	4	4	3	3	10	23,46
3		5	4	4	8,57	9,88
2					0	0
5					0	2,47
					18,57	35,81
4	4	5	3	3	2,86	11,11
3			4	4	1,43	4,94
2					0	0
5					0	2,47
					4,29	18,52

Pourcentage d'apparition des profils reprenant des notes extrêmes en langue maternelle.

(1) En sciences, la fréquence d'apparition des pourcentages de tendance centrale est quasi nulle dans les deux groupes et lors des deux expérimentations, en raison des contrastes accusés existant dans le travail analysé.

Sciences

AG	Desc.	Int.	Sch.	Inv.	Gr. II	Gr. III
4 3 2 5	5	3 2 1	4 5	3 1	3,60	13,27
					7,21 0,90 0	4,42 0,88 0
					11,71	12,57
4 3 2 5	4 5	3 2 1	4	3 1	0,90	7,08
					10,81 1,80 0	6,19 1,77 0
					13,51	14,94
4 3 2 5	5	3 2 1	4	3 1	0	4,42
					4,50 0,90 0	0 0 0
					5,40	4,42

Pourcentages d'apparition des profils utilisant des notes extrêmes en sciences.

b) Les échelles descriptives aident-elle

1. — Analyse des corrélations

Langue maternelle.

	PS
Pertinence sémantique	1
Pertinence syntaxique	
Vocabulaire	
Structures grammaticales	
Organisation des idées	

	PS
Pertinence sémantique	1
Pertinence syntaxique	
Vocabulaire	
Structures grammaticales	
Organisation des idées	

Gr. II	Gr. III
3,60	13,27
7,21 0,90 0	4,42 0,88 0
11,71	12,57
0,90	7,08
10,81 1,80 0	6,19 1,77 0
13,51	14,94
0	4,42
4,50 0,90 0	0 0 0
5,40	4,42

t des notes extrêmes en sciences.

b) Les échelles descriptives aident-elles à lutter contre l'effet de contamination ?

1. — Analyse des corrélations entre appréciations analytiques.

Langue maternelle.

	PS	Ps	Voc.	SG	Org.
Pertinence sémantique	1	.12 N.S.	.36 S à P.005	0 N.S.	.40 S à P.005
Pertinence syntaxique		1	.10 N.S.	0 N.S.	0.33 S à P.005
Vocabulaire			1		.24 S à P.025
Structures grammaticales				1	.07 N.S.
Organisation des idées					1

Groupe II
(n = 70)

	PS	Ps	Voc.	SG	Org.
Pertinence sémantique	1	.03 N.S.	.17 N.S.	.15 N.S.	.15 N.S.
Pertinence syntaxique		1	.38 S à P.005	.28 S à P.005	.13 N.S.
Vocabulaire			1	.14 N.S.	.21 S à P.025
Structures grammaticales				1	.10 N.S.
Organisation des idées					1

Groupe III
(n = 81)

Sciences

	Des.	Sch.	Int.	Inv.
Description	1	.40 S à P.005	.14 N.S.	.05 N.S.
Schéma		1	.14 N.S.	.00 N.S.
Interprétation			1	.36 S à P.005
Invention				1

Groupe II (n = 111)

	Des.	Sch.	Int.	Inv.
Description	1	.54 S à P.005	.14 N.S.	.06 N.S.
Schéma		1	.25 S à P.01	.09 N.S.
Interprétation			1	.21 S à P.02
Invention				1

Groupe III (n = 113)

2. — Analyse factorielle des notes analytiques.

Langue maternelle

	Gr. II			Gr. III		
	F1	F2	F3	F1	F2	F3
% cumulatif des valeurs propres	.35	.59	.77	.34	.55	.73
<i>Saturation de chaque variable</i>						
Pertinence sémantique.....	-.04	.83	.09	.42	.71	.35
Pertinence syntaxique	-.02	.01	.94	.69	-.50	-.6
Vocabulaire05	.80	.04	.71	-.9	-.34
Structures grammaticales95	-.04	.02	.56	-.18	.66
Organisation des idées47	.34	.58	.50	.44	-.45

Sciences

% cumulatif des valeurs propres	
Description	
Schéma	
Interprétation	
Invention	

4. — UNE ÉVALUATION OBJECTIVE N

a) Pour évaluer, les juges n'ont-ils leur proposait ?

La part prise par les notes anal le pourcentage de variance expliqu

Langue maternelle

	% de
Gr. II	
Gr. III	

Sciences

	% de
Gr. II	
Gr. III	

(*) La part de variance inexplic à des facteurs aléatoires (facteurs tairement par les professeurs (facteu tinguer l'influence respective de ce

	Int.	Inv.
	.14 N.S.	.05 N.S.
	.14 N.S.	.00 N.S.
	1	.36 S à P.005
		1

	Int.	Inv.
	.14 N.S.	.06 N.S.
	.25 S à P.01	.09 N.S.
	1	.21 S à P.02
		1

F3	Gr. III		
	F1	F2	F3
.77	.34	.55	.73
.09	.42	.71	.35
.94	.69	-.50	-.6
.04	.71	-.9	-.34
.02	.56	-.18	.66
.58	.50	.44	-.45

Sciences

	Gr. II		Gr. III	
	F1	F2	F1	F2
% cumulatif des valeurs propres .	.39	.69	.42	.69
Description82	.08	.87	.01
Schéma84	.01	.86	.15
Interprétation19	.79	.24	.70
Invention	-.01	.84	-.05	.83

4. — UNE ÉVALUATION OBJECTIVE N'EST PAS INFLUENCÉE PAR DES FACTEURS INTERNES.

a) Pour évaluer, les juges n'ont-ils pas fait intervenir d'autres critères que ceux qu'on leur proposait ?

La part prise par les notes analytiques dans la note de synthèse s'apprécie par le pourcentage de variance expliquée (*) et par le coefficient de corrélation multiple.

Langue maternelle

	% de variance expliquée	Coefficient de corrélation multiple
Gr. II	46 %	.681
Gr. III	45 %	.651

Sciences

	% de variance expliquée	Coefficient de corrélation multiple
Gr. II	60 %	.774
Gr. III	30 %	.536

(*) La part de variance inexpliquée par les dimensions analytiques peut être due à des facteurs aléatoires (facteurs externes) ou à d'autres critères introduits volontairement par les professeurs (facteurs internes). La technique ne permet pas de distinguer l'influence respective de ces deux sources de variance.

HYPOTHÈSE 2 : L'ÉCHELLE DESCRIPTIVE PERMET UNE ÉVALUATION PLUS VALIDE

1. — LA NOTE GLOBALE EST-ELLE CORRÉLÉE SIGNIFICATIVEMENT AVEC CHACUNE DES APPRÉCIATIONS ANALYTIQUES ?

a) *Corrélation de la note globale avec chaque appréciation analytique. Analyse des corrélations simples.*

Langue maternelle

	AG	PS	Ps	Voc.	S.G.	Org.
AG	1	.58 S à P.005	.18 N.S.	.50 S à P.005	.06 N.S.	.40 S à P.005

	AG	PS	Ps	Voc.	S.G.	Org.
AG	1	.30	.34 ← S à P.005 →	.50	.24	.41

Sciences

Groupe II

	AG	Des.	Sch.	Int.	Inv.
AG	1	.53 ← S à P.005 →	.51	.52	.37

Groupe III

	AG	Desc.	Sch.	Int.	Inv.
AG	1	.22	.36 ← S à P.005 →	.42	.34

b) *Analyse factorielle.*

En introduisant la note globale, quelles variables, saturées dans la dernière, mieux cette dernière.

Langue maternelle

% cumulatif des valeurs propres
Appréciation globale
Pertinence sémantique
Pertinence syntaxique
Vocabulaire
Structures grammaticales
Organisation des idées

Sciences

% cumulatif des valeurs propres
Appréciation globale
Description
Schéma
Interprétation
Invention

c) *Analyse par segmentation.*

Cette technique permet de recueillir une population initiale en deux sous-populations (à partir d'une appréciation globale). La meilleure coupe est celle qui minimise la distance entre les moyennes des deux sous-populations.

On présente successivement l'analyse sous forme de diagramme de Venet (au moins ont été reprises dans les domaines délimités sont proportionnelles aux valeurs du prédicteur utilisées).

RIPTIVE PERMET UNE
S VALIDE

EFICACEMENT AVEC CHACUNE DES

ppréciation analytique. Analyse des

Voc.	S.G.	Org.
.50	.06	.40
S à P.005	N.S.	S à P.005

Voc.	S.G.	Org.
.50	.24	.41
P.005 →		

h.	Int.	Inv.
1	.52	.37

h.	Int.	Inv.
6	.42	.34
S à P.005 →		

b) *Analyse factorielle.*

En introduisant la note globale dans l'analyse factorielle, on cherche à déterminer quelles variables, saturées dans le même facteur que la note globale expliquent au mieux cette dernière.

Langue maternelle

	Groupe II			Groupe III		
	F1	F2	F3	F1	F2	F3
% cumulatif des valeurs propres	.38	.59	.74	.38	.55	.70
Appréciation globale85	-.01	.14	.82	-.14	-.12
Pertinence sémantique82	-.02	.07	.43	-.59	.50
Pertinence syntaxique03	-.02	.94	.61	.58	-.8
Vocabulaire74	.08	.03	.7	.12	-.30
Structures grammaticales	-.04	.95	.01	.47	.34	.67
Organisation des idées31	.48	.57	.54	-.42	-.30

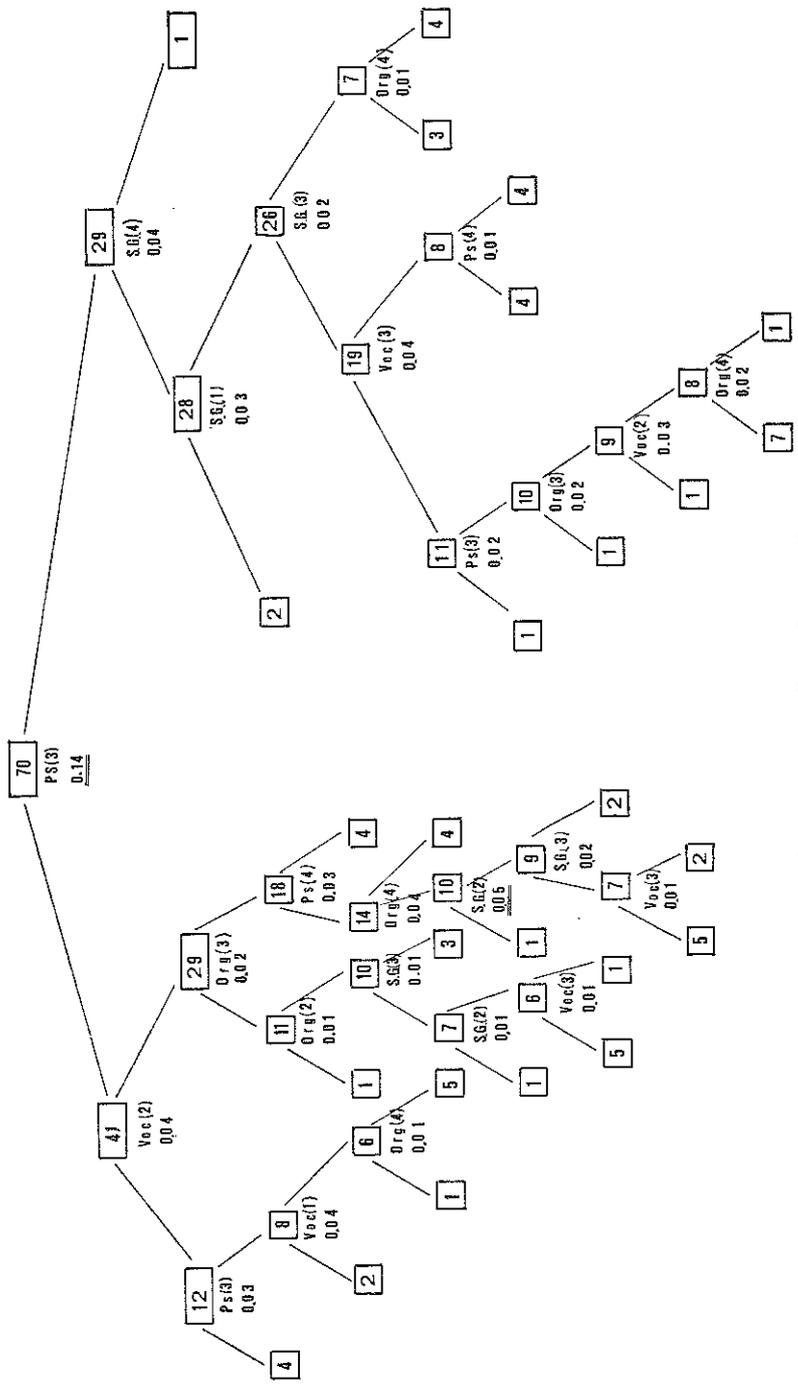
Sciences

	Groupe II		Groupe III	
	F1	F2	F1	F2
% cumulatif des valeurs propres	.46	.70	.42	.65
Appréciation globale70	.57	.74	.30
Description80	.05	.02	.87
Schéma82	-.01	.20	.84
Interprétation19	.78	.69	.18
Invention	-.07	.83	.74	-.11

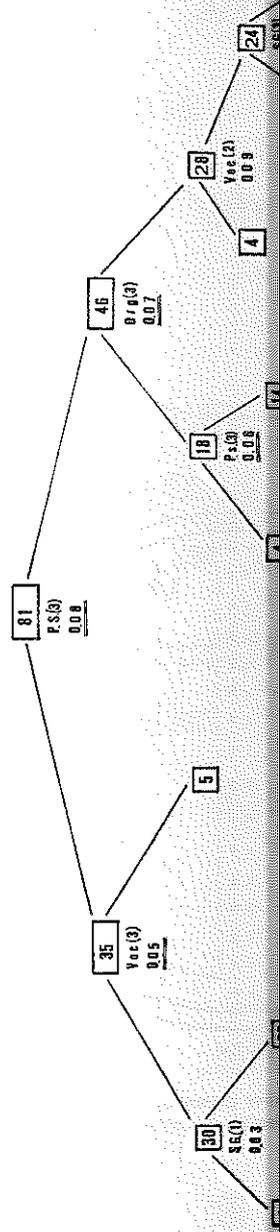
c) *Analyse par segmentation.*

Cette technique permet de rechercher quel prédicteur — ici quelle appréciation analytique — et quelle valeur de ce prédicteur — ici valeur de 1 à 5 — scinde au mieux une population initiale en deux sous-populations pour un critère donné (ici l'appréciation globale). La meilleure coupure de la population initiale est celle qui maximise la distance entre les moyennes des deux sous-populations.

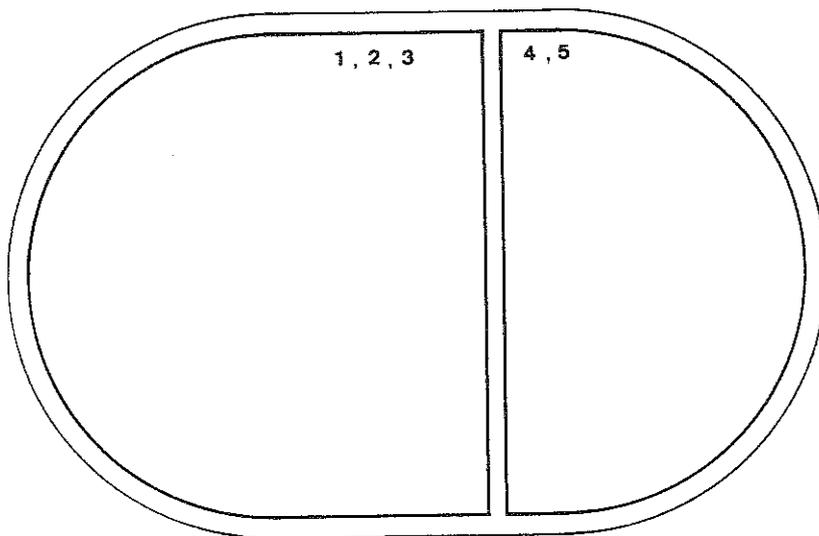
On présente successivement l'arbre de segmentation et sa représentation graphique sous forme de diagramme de Venn. Seules les sous-populations distantes de 0,05 point au moins ont été reprises dans le diagramme. Pour chaque sous-population, les domaines délimités sont proportionnels au nombre de sujets. On indique en outre les valeurs du prédicteur utilisées dans chacune des sous-populations.



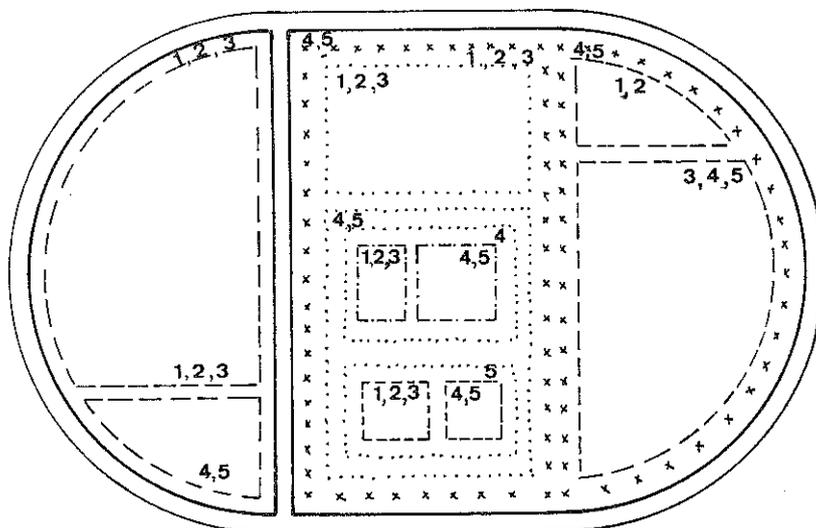
Langue maternelle - Groupe II



Langue maternelle

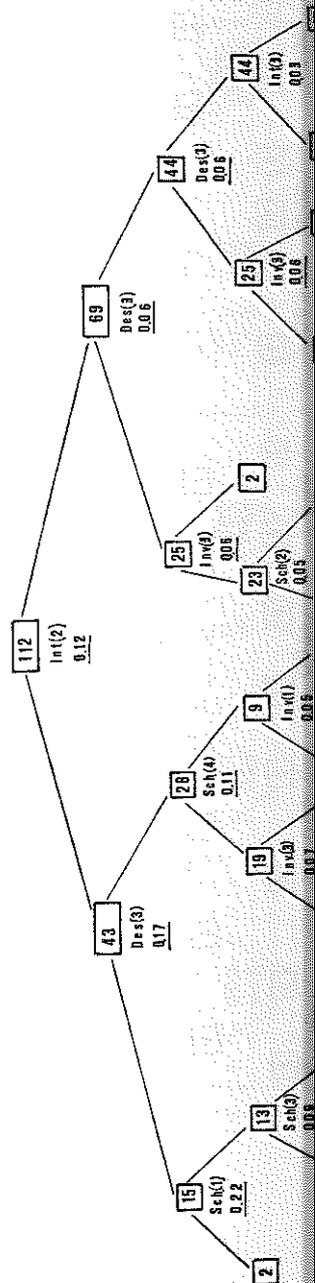


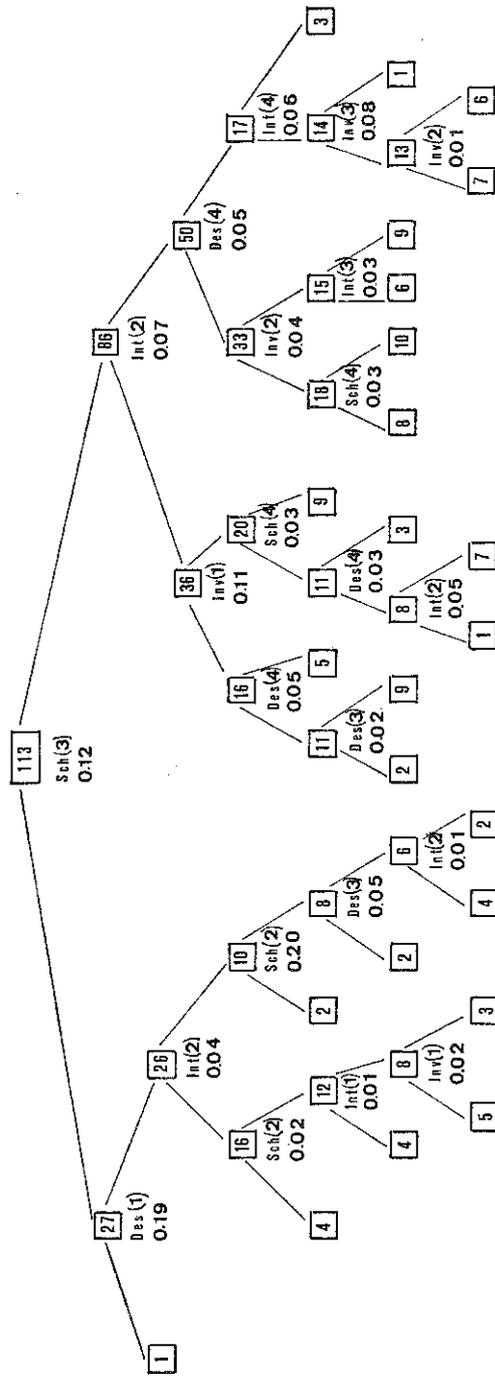
Groupe II



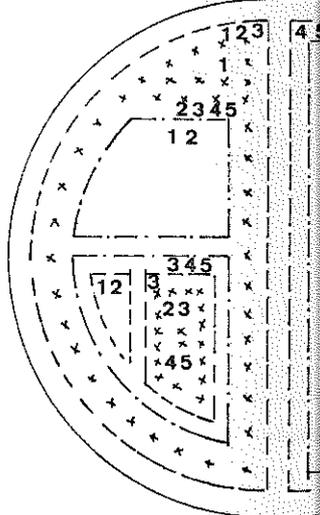
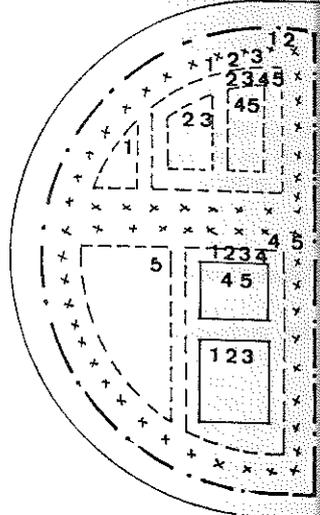
Groupe III

- Pertinence sémantique
- Pertinence syntaxique
- Vocabulaire
- Structures grammaticales
- xxxxx Organisation des idées

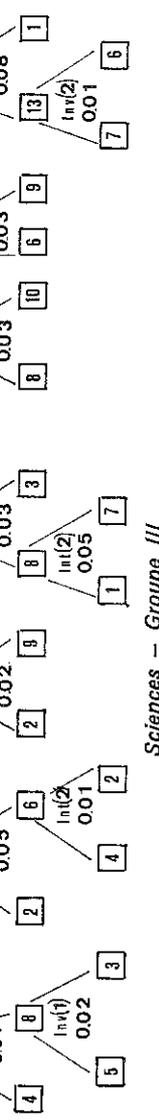




Sciences - Groupe III

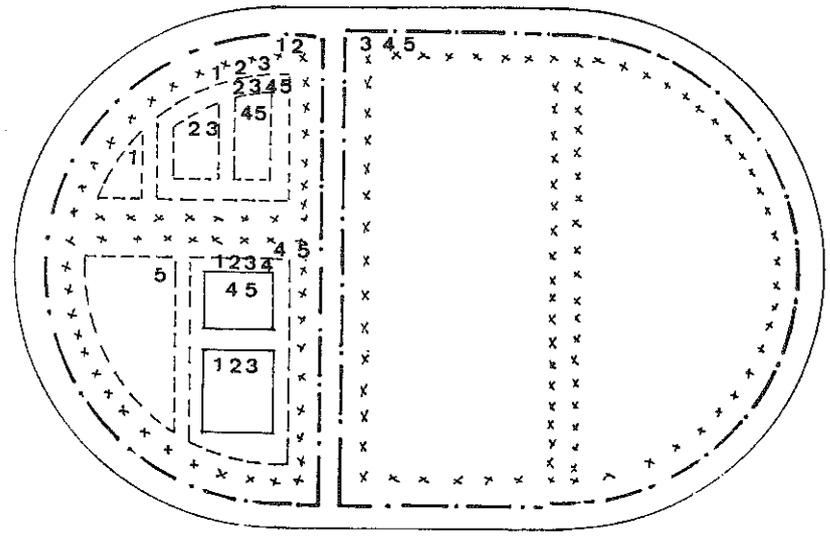


— Invention
 - - - Schéma

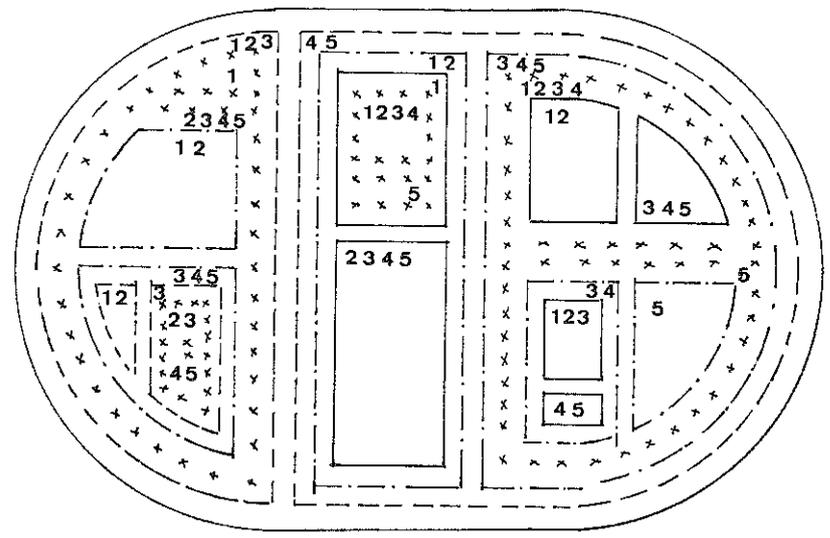


Sciences - Groupe III

Sciences



Groupe II



Groupe III

——— Invention
 - - - - Schéma
 - - - - Interprétation
 x x x x Description

HYPOTHÈSE 3 : L'ÉCHELLE DESCRIPTIVE PERMET UNE ÉVALUATION MIEUX UTILISABLE PAR LE PROFESSEUR ET PAR L'ÉLÈVE

Le tableau 1 présente la manière dont les différentes justifications sont utilisées. On lira successivement, pour chaque groupe, le nombre de cas où aucune justification n'est donnée (1^{re} colonne), le nombre de fois où elle est positive (2^e colonne) et le nombre de fois où elle est négative (3^e colonne).

Les justifications sont réparties en trois catégories :

- 1° Justifications correspondant aux critères faisant l'objet d'échelles (justifications principales).
- 2° Justifications apparues dans les commentaires des professeurs (justifications secondaires).
- 3° Justifications inclassables par manque de précision.

Le tableau 2 présente les mêmes informations traduites en pourcentages.

Langue maternelle

Justifications principales	Pertinence sémantique Pertinence syntaxique Vocabulaire Structures grammaticales Organisation des idées
	Total
Justifications secondaires	Orthographe Ponctuation Ton Imagination Congugaison
Total général des justifications classées	
Justifications inclassables	

+ : positives - : négatives

TIVE PERMET UNE
UTILISABLE PAR LE
L'ÉLÈVE

es justifications sont utilisées.
de cas où aucune justification
est positive (2^e colonne) et le

objet d'échelles (justifications

es professeurs (justifications

ites en pourcentages.

TABLEAU I. *Fréquences brutes*

Langue maternelle

		Groupe I (n = 101)				Groupe II (n = 70)				Groupe III (n = 81)			
		Néant	Justification			Néant	Justification			Néant	Justification		
			+	-	Tot a+b		+	-	Tot a+b		+	-	Tot a+b
Justifications principales	Pertinence sémantique	46	27	28	55	58	5	7	12	60	15	6	21
	Pertinence syntaxique	99	0	2	2	68	0	2	2	62	2	17	19
	Vocabulaire	75	9	17	26	57	0	13	13	54	17	10	27
	Structures grammaticales	52	22	27	49	55	3	12	15	56	17	8	25
	Organisation des idées	70	16	15	31	63	6	1	7	60	2	19	21
	Total		74	89	163		14	35	49		53	60	113
Justifications secondaires	Orthographe	31	0	70	70	53	1	16	17	60	2	19	21
	Ponctuation	45	3	53	56	63	0	7	7	75	6	0	6
	Ton	85	7	9	16	63	6	1	7	78	2	1	3
	Imagination	77	11	13	24	58	0	12	12	72	9	0	9
	Congugaison	91	5	5	10	68	1	1	2	58	4	19	23
		26	150	176		8	37	45		23	39	62	
Total général des justifications classées			100	239	339		22	72	94				175
Justifications inclassables					84				45				64

+ : positives - : négatives

TABLEAU I. *Fréquences brutes*

Sciences

		Groupe I (n = 122)				Groupe II (n = 111)				Groupe III (n = 113)			
		Néant	Justification			Néant	Justification			Néant	Justification		
			+	-	Tot		+	-	Tot		+	-	Tot
		a	b	a+b	a	b	a+b	a	b	a+b			
Justifications principales	Description	74	41	7	48	95	15	1	16	86	22	5	27
	Schéma	70	50	2	52	99	11	1	12	89	22	2	24
	Interprétation	57	4	61	65	92	6	13	19	84	4	25	29
	Invention	46	3	73	76	99	0	12	12	98	0	15	15
	Total		98	143	241		32	27	59		48	47	95
Justifications secondaires	Littérature	113	2	7	9	104	3	2	7	105	3	5	8
	Observation	106	15	1	16	97	13	1	14	93	20	0	20
	Conclusion	104	1	17	18	104	1	6	7	107	0	6	6
	Compréhension objet de l'expér.	93	7	22	29	93	8	10	18	97	6	10	16
	Distinction Description-Interprétation	110	1	11	12	107	0	4	4	104	0	4	4
Total		26	58	84		25	25	50		29	25	54	
Total des justific. classées			124	201	325		57	52	109		77	72	149

+ : positives - : négatives

Langue maternelle

Justifications principales	Pertinence sémantique Pertinence syntaxique Vocabulaire Structures grammaticales Organisation des idées
Justification secondaire	Orthographe Ponctuation Ton. Imagination Conjugaison

Sciences

Justific. princip.	Description Schéma Interprétation Invention
Justifications secondaires	Littérature Observation Conclusion Compréhension objet de l'expérience Distinction Description-Interprétation

TABLEAU II. Pourcentages

Langue maternelle

Groupe III (n = 113)		Justification		
Justification	Néant	+	-	Tot a+b
		a	b	
16	86	22	5	27
12	89	22	2	24
19	84	4	25	29
12	98	0	15	15
59		48	47	95
7	105	3	5	8
14	93	20	0	20
7	107	0	6	6
18	97	6	10	16
4	104	0	4	4
50		29	25	54
109		77	72	149

		Groupe I			Groupe II			Groupe III		
		Néant	Justifications		Néant	Présente		Néant	Présente	
			Positives a	Négatives b		Positives a	Négatives b		Positives a	Négatives b
Justifications principales	Pertinence sémantique	46	27	28	83	7	10	74	18	9
	Pertinence syntaxique	98	0	2	97	0	3	97	2	0
	Vocabulaire	74	9	17	81	0	19	76	2	20
	Structures grammaticales	51	22	27	79	4	17	66	20	12
	Organisation des idées	69	16	15	90	9	1	69	20	9
Justifications secondaires	Orthographe	31	0	69	76	1	23	74	1	23
	Ponctuation	45	3	52	90	0	10	74	2	23
	Ton.	84	7	9	90	9	1	92	7	0
	Imagination	76	11	13	83	0	17	89	11	1
	Conjugaison	90	5	5	97	1	1	96	2	1

Sciences

Justific. princip.	Description	61	33	6	86	13	1	76	19	5
		57	41	2	89	10	1	79	19	2
Justifications secondaires	Schéma	47	3	50	83	5	12	74	4	22
	Interprétation	38	2	60	89		11	87	—	13
	Invention	93	2	5	94	3	3	93	3	4
	Littérature	87	12	1	87	12	1	82	18	5
Justifications secondaires	Observation	85	1	14	94	1	5	95	—	5
	Conclusion	76	6	18	84	7	9	86	5	9
	Compréhension objet de l'expérience	90	1	9	96	—	4	96	0	4
	Distinction									
	Description-Interprétation									

1. — LE NOMBRE DE JUSTIFICATIONS DIFFÈRE-T-IL D'UN GROUPE À L'AUTRE ?

a) *Nombre global des justifications.*

	Groupe I	Groupe II	Groupe III
Nombre total de justifications .	339	94	175
Nombre de professeurs	101	70	81
Nombre moyen par professeur	3,3	1,3	2,1

b) *Répartition du nombre de justifications utilisées par chaque professeur.*

Langue maternelle

Les tableaux et graphiques suivants reprennent le nombre de justifications par professeur.

	I	%	II	%	III	%
0 justification	1	0,9	13	19,6	24	29,2
1 justification	8	7,9	30	45,4	17	20,7
2 justifications	24	34,6	14	21,2	20	24,4
3 justifications	35	25,7	6	9,1	11	13,4
4 justifications	26	6,9	3	4,5	7	8,5
5 justifications	7	—	—	—	2	2,4
6 justifications	—	—	—	—	—	—
7 justifications	—	—	—	—	1	1,2

2. — LES PROFESSEURS UTILISENT-ILS DES JUSTIFICATIONS EN RAPPORT AVEC LES OBJECTIFS DU TRAVAIL PROPOSÉ ?

D'abord, quels critères les professeurs utilisent-ils spontanément lorsqu'ils évaluent librement ?

On reprend, dans le tableau de la page 235, les données relatives au groupe I.

Langue maternelle

Justifications principales
Justifications secondaires
Total des justifications classées

Sciences

Justifications principales
Justifications secondaires
Total des justifications classées

a) *Proportion des justifications différents groupes.*

Sur l'ensemble des justifications des justifications principales (et des justifications secondaires).

GRUPE À L'AUTRE ?

Groupe II	Groupe III
94	175
70	81
1,3	2,1

pour chaque professeur.

nombre des justifications par

%	III	%
19,6	24	29,2
45,4	17	20,7
21,2	20	24,4
9,1	11	13,4
4,5	7	8,5
	2	2,4
	1	1,2

EN RAPPORT AVEC LES OBJECTIFS

spontanément lorsqu'ils évaluent

années relatives au groupe I.

Langue maternelle

		Fréquences brutes	%
Justifications principales	Pertinence sémantique	55	16,2
	Pertinence syntaxique	2	0,6
	Vocabulaire	26	7,7
	Structures grammaticales	49	14,5
	Organisation des idées	31	9,1
Justifications secondaires	Orthographe	70	20,6
	Ponctuation	56	16,5
	Ton	16	4,7
	Imagination	24	7,1
	Conjugaison	10	2,9
Total des justifications classées		339	100 %

Sciences

		Fréquences brutes	%
Justifications principales	Description	48	14,8
	Schéma	52	16
	Interprétation	65	20
	Invention	76	23,4
Justifications secondaires	Littérature	9	2,8
	Observation	16	4,9
	Conclusion	18	5,5
	Compréhension de l'objet de l'expérience	29	8,9
	Distinction		
	Description-interprétation	12	3,7
Total des justifications classées		325	100 %

a) Proportion des justifications principales et des justifications secondaires dans les différents groupes.

Sur l'ensemble des justifications positives ou négatives, on calcule la proportion des justifications principales (en relation avec les objectifs du travail) et des justifications secondaires.

Langue maternelle

Justifications	Groupe I	Groupe II	Groupe III
Principales	163/339 = 48 %	49/94 = 52 %	113/175 = 65 %
Secondaires	176/339 = 52 %	45/94 = 48 %	62/175 = 35 %

La comparaison des pourcentages montre que la proportion des justifications principales par rapport aux justifications secondaires est nettement plus importante dans le groupe III. Cette répartition est due au poids que prennent l'orthographe et la ponctuation dans le groupe I. Encore faut-il tester la signification statistique des différences observées (1).

	I	II	III	
Principales	163 (2) (181)	49 (50)	113 (94)	325
Secondaires	176 (158)	45 (44)	62 (81)	283
	339	94	175	608

$\chi^2 = 12,17$ S à P.01

	I	II	
Principales	163 (166)	49 (46)	212
Secondaires	176 (173)	45 (48)	221
	339	94	433

$\chi^2 = 0,65$ N.S.

	II	III	
Principales	49 (57)	113 (105)	162
Secondaires	45 (37)	62 (70)	107
	94	175	269

$\chi^2 = 4,36$ S à P.05

(1) Le test du χ^2 permet de contrôler la signification des différences entre les fréquences.

(2) Les nombres entre parenthèses représentent les fréquences théoriques.

I III

Principales	163 (182)	113 (94)
Secondaires	176 (157)	62 (81)
	339	175

La différence significative entre l'opposition entre le groupe III et les autres groupes est statistiquement équivalente.

Sciences

Justifications principales	241
Justifications secondaires	84

La comparaison des pourcentages principales opposées aux secondaires

	I	II
Principales	241 (220)	50 (50)
Secondaires	84 (105)	50 (50)
	325	100

	I	II
Principales	241 (225)	50 (50)
Secondaires	84 (100)	50 (50)
	325	100

Groupe II	Groupe III
= 52 %	113/175 = 65 %
= 48 %	62/175 = 35 %

proportion des justifications principales plus importante dans le groupe I et II que dans le groupe III. La différence est statistiquement significative des différences théoriques.

$\chi^2 = 12,17$ S à P.01

$\chi^2 = 0,65$ N.S.

$\chi^2 = 4,36$ S à P.05

différences entre les fréquences théoriques.

	I	III	
Principales	163 (182)	113 (94)	276
Secondaires	176 (157)	62 (81)	238
	339	175	514

$\chi^2 = 12,57$ S à P.001

La différence significative enregistrée pour l'ensemble des répartitions naît de l'opposition entre le groupe III et les groupes I et II dont les pourcentages sont pratiquement équivalents.

Sciences

	Groupe I	Groupe II	Groupe III
Justifications principales	241/325 = 74 %	59/109 = 54 %	95/149 = 64 %
Justifications secondaires	84/325 = 26 %	50/109 = 46 %	54/149 = 36 %

La comparaison des pourcentages montre que la proportion des justifications principales opposées aux secondaires est plus importante dans le groupe I.

	I	II	III	
Principales	241 (220)	59 (74)	95 (101)	395
Secondaires	84 (105)	50 (35)	54 (48)	188
	325	109	149	583

$\chi^2 = 16,78$ S à P.001

	I	II	
Principales	241 (225)	59 (75)	300
Secondaires	84 (100)	50 (34)	134
	325	109	434

$\chi^2 = 14,64$ S à P.001

	II	III	
Principales	59 (65)	95 (89)	154
Secondaires	50 (44)	54 (60)	104
	109	149	258

$\chi^2 = 2,37$ N.S.

	I	III	
Principales	241 (230)	95 (106)	336
Secondaires	84 (95)	54 (43)	138
	325	149	474

$\chi^2 = 8,12$ S à P.01

b) *Caractéristiques des justifications par professeur.*

Langue maternelle

	I		II		III	
	Fréq. brutes	%	Fréq. brutes	%	Fréq. brutes	%
Pas de justification	1	0,9	13	19,6	24	29,2
Justif. second. seulement	20	19,8	17	25,7	21	25,6
Seulement 1 justif. princ., quelque soit le nombre de justif. second.	51	50,50	12	18,18	27	32,9
2 justif. principales	22	21,78	10	15,15	14	17,1
3 justif. principales	7	6,93	2	3,03	3	3,7
4 justif. principales	0	—	1	1,52	4	4,9

Sciences

	Fréq. brutes
Pas de justification	1
Justif. second. uniq ^t	14
1 principale seulement	21
2 principales seulement	21
3 principales seulement	11
4 principales seulement	2

c) *Importance respective de chaque justification principale.*

Langue maternelle

Total des justifications principales	10
Pertinence sémantique	5
Pertinence syntaxique	2
Vocabulaire	4
Structures grammaticales	3
Organisation des idées	3

Sciences

Total des justifications principales	2
Description	4
Schéma	1
Interprétation	1
Invention	1

Sciences

	I		II		III	
	Fréq. brutes	%	Fréq. brutes	%	Fréq. brutes	%
Pas de justification	9	7,38	26	23,4	31	27,4
Justif. second. uniq ^t	14	11,48	19	17,1	27	23,8
1 principale seulement	28	22,9	10	9	28	24,7
2 principales seulement	27	22,1	10	9	14	12,3
3 principales seulement	19	15,5	4	3,6	11	9,7
4 principales seulement	25	20,4	1	0,9	2	1,7

$\chi^2 = 2,37$ N.S.

$\chi^2 = 8,12$ S à P.01

c) Importance respective de chaque justification principale dans l'ensemble des justifications principales.

Langue maternelle

II	III	
	Fréq. brutes	%
19,6	24	29,2
25,7	21	25,6
18,18	27	32,9
15,15	14	17,1
3,03	3	3,7
1,52	4	4,9

Total des justifications principales	Groupe I		Groupe II		Groupe III	
	n	%	n	%	n	%
	163	100	49	100	113	100
Pertinence sémantique	55	33,71	12	24,5	21	18,6
Pertinence syntaxique	2	1,2	2	4,1	19	16,8
Vocabulaire	26	15,9	13	26,5	27	23,9
Structures grammaticales	49	30,1	15	30,6	25	22,1
Organisation des idées	31	19,1	7	14,2	21	18,5

Sciences

Total des justifications principales	Groupe I		Groupe II		Groupe III	
	n	%	n	%	n	%
	241	100	59	100	95	100
Description	48	20	16	27	27	22
Schéma	52	22	12	20	24	20
Interprétation	65	27	19	32	29	22
Invention	76	32	12	20	15	12

3. — LES COMMENTAIRES DESTINÉS À L'ÉLÈVE VISENT-ILS DES ASPECTS FACILEMENT IDENTIFIABLES PAR CE DERNIER ?

a) *Les justifications des professeurs fournissent-elles à l'élève des informations assez précises pour qu'il puisse en tirer concrètement parti dans la suite de son apprentissage?*

	Groupe I	Groupe II	Groupe III
Nombre total des justifications classées	339	94	175
Nombre de justifications spécifiques	44	4	26
Pourcentage	12,9	4,2	14,8

b) *Observe-t-on de nombreuses justifications portant sur des détails du travail de l'élève, sur des caractéristiques marginales ?*

Ce type de justifications apparaît principalement dans l'évaluation de la pertinence sémantique. Par exemple, certains professeurs mentionnent uniquement un détail oublié par l'élève dans la rédaction du paragraphe intermédiaire (la coupure, la porte refermée,...)

	Groupe I	Groupe II	Groupe III
Nombre total des justifications classées	339	94	175
Nombre de justifications s'attachant aux détails	24	1	4
Pourcentage	7	1	2

c) *Enfin, quelle est la répartition des justifications « inclassables » ?*

	Groupe I	Groupe II	Groupe III
Nombre total des justifications	423	139	237
Nombre de justifications inclassables	84	45	64
Pourcentage	19,8	32,3	27

Parmi les justifications que l'on trouve, certaines sont suffisamment spécifiques pour qu'on puisse en tirer parti dans l'apprentissage. Certaines s'attachent à des détails oubliés. Le pourcentage des justifications...

4. — LES PROFESSEURS RELÈVENT-ILS DES ASPECTS POSITIFS OU NÉGATIFS DU TRAVAIL ?

a) *Analyse de l'ensemble des justifications*

Langue maternelle

Justifications principales ...	4
Pourcentage	2
Justifications secondaires ...	1
Pourcentage	0,5
Total des justif. classées ...	5
Pourcentage	2,5

Sciences

Justifications principales ...	4
Pourcentage	2
Justifications secondaires ...	1
Pourcentage	0,5
Total des justif. classées ...	5
Pourcentage	2,5

ILS DES ASPECTS FACILEMENT

*l'élève des informations assez
is la suite de son apprentissage?*

Groupe II	Groupe III
94	175
4	26
4,2	14,8

les détails du travail de l'élève,

*s l'évaluation de la pertinence
onnent uniquement un détail
médiaire (la coupure, la porte*

Groupe II	Groupe III
94	175
1	4
1	2

classables » ?

Groupe II	Groupe III
139	237
45	64
32,3	27

Parmi les justifications que l'on peut classer dans une catégorie définie, peu sont suffisamment spécifiques pour que l'élève puisse en tirer parti pour la suite de son apprentissage. Certaines s'attachent uniquement à des détails sans grande signification. Le pourcentage des justifications inclassables est relativement élevé.

4. — LES PROFESSEURS RELÈVENT-ILS DANS LEURS COMMENTAIRES, LES ASPECTS POSITIFS OU NÉGATIFS DU TRAVAIL ?

a) *Analyse de l'ensemble des justifications.*

Langue maternelle

	I			II			III		
	Pos.	Nég.	Tot.	Pos.	Nég.	Tot.	Pos.	Nég.	Tot.
Justifications principales ...	74	89	163	14	35	49	53	60	113
Pourcentage	45,3	54,6		28,5	71,4		46,9	53,1	
Justifications secondaires ..	26	150	176	8	37	45	23	39	62
Pourcentage	14,7	85,2		17,7	82,2		37,1	62,9	
Total des justif. classées ...	100	239	339	22	72	94	76	99	175
Pourcentage	28,4	70,5		23,4	76,5		43,4	56,5	

Sciences

	I			II			III		
	Pos.	Nég.	Tot.	Pos.	Nég.	Tot.	Pos.	Nég.	Tot.
Justifications principales ...	98	143	241	32	27	59	48	47	95
Pourcentage	40,6	59,3		54,2	45,7		50,5	49,5	
Justifications secondaires ..	26	58	84	25	25	50	29	25	54
Pourcentage	31,0	69,0		50,0	50,0		53,7	4,3	
Total des justif. classées ...	124	201	325	57	52	109	77	72	149
Pourcentage	38,2	61,8		52,3	47,7		51,7	48,3	

b) Analyse des justifications principales.

Langue maternelle

	Gr. I		Gr. II		Gr. III	
	positives nb. % brut	négatives nb. % brut	positives nb. % brut	négatives nb. % brut	positives nb. % brut	négatives nb. % brut
Total des justif. princ...	74 45	28 51	14 29	35 71	53 49	60 53
P. sémantique	27 36,4	28 31,4	5 35,7	7 20	15 28,3	6 10
P. syntaxique	0 0	2 2,2	0 —	2 5,7	2 3,7	17 28,3
Vocabulaire	9 12,1	17 19,1	0 —	13 37,1	17 32,1	10 16,7
Struct. grammat.	22 29,7	27 30,3	3 21,4	12 34,2	17 32,1	18 30
Organisation des idées	16 21,6	15 16,8	6 42,8	1 2,8	2 3,7	19 31,6

Sciences

	Gr. I		Gr. II		Gr. III	
	positives nb. % brut	négatives nb. % brut	positives nb. % brut	négatives nb. % brut	positives nb. % brut	négatives nb. % brut
Total des justif. princ...	98 42	143 59	32 54	27 46	48 50	47 48,5
Description	41 42	7 5	15 47	1 4	22 46	5 11
Schéma	50 51	2 1	11 34	1 4	22 46	2 4
Interprétation	4 4	61 43	6 19	13 48	4 8	25 53
Invention	3 3	13 51	0 —	12 44	0 —	15 32