

Introduction aux sciences de l'éducation et de la formation (ISEF)

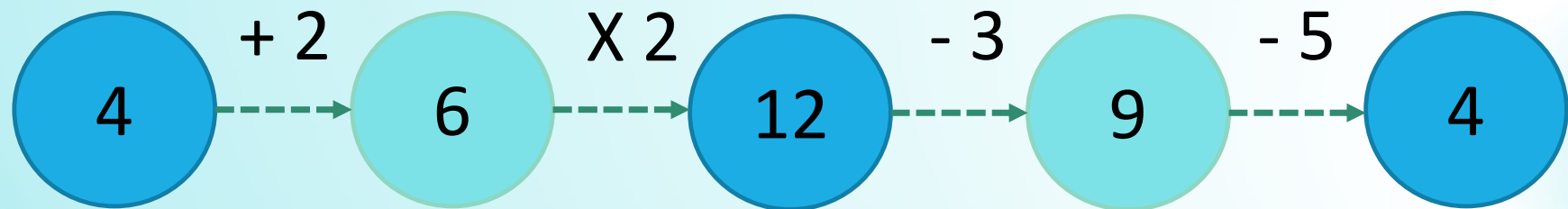
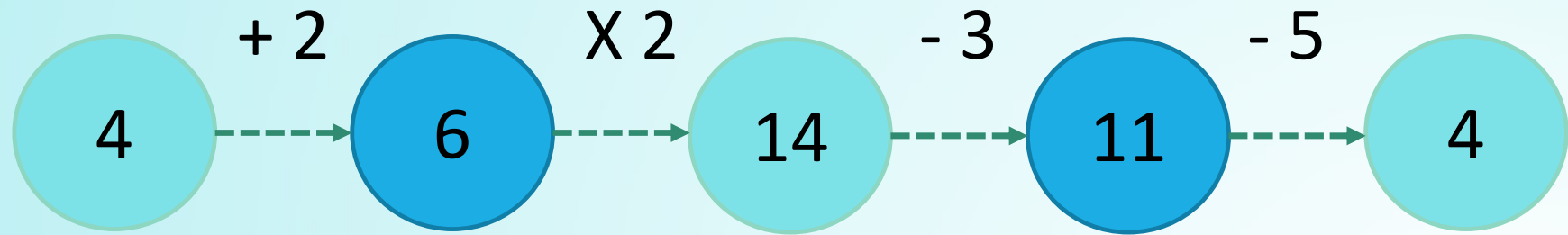
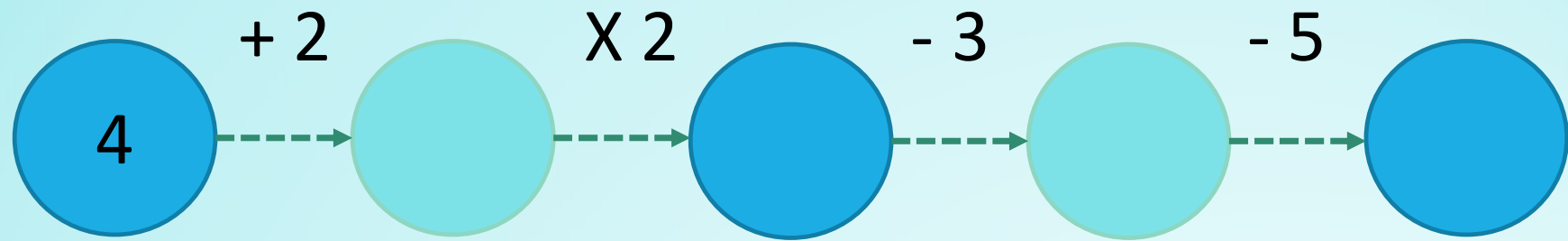
PEDA 4021

14 novembre 2024

Prof. : Ariane Baye, aide@uliege.be

Chapitre 6

l'évaluation des élèves : Docteur Jekyll and Mister Hyde de l'enseignement



Quelle note pour le calcul en cascade?

- 1 pour 0/2
- 2 pour 0,5 / 2
- 3 pour 1/2
- 4 pour 1,5/2
- 5 pour 2/2

Une évaluation objective est-elle possible ?

Plan du cours

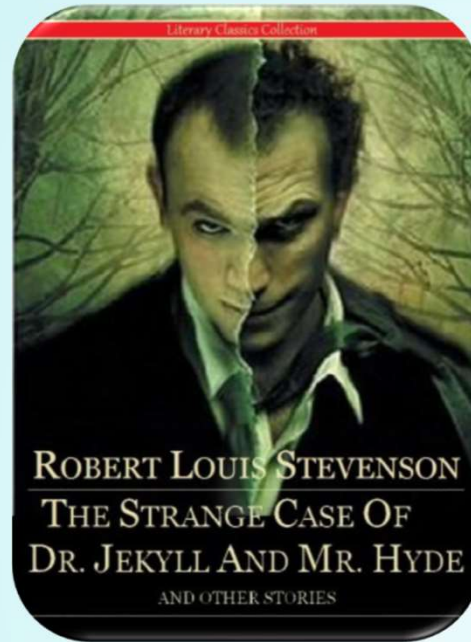
1. Introduction, des fonctions de l'évaluation ?
2. Des fonctions multiples, mais une visée de régulation
3. Confusions entre les fonctions de l'évaluation
4. Évaluer dans une perspective d'excellence ou de maîtrise ?
5. Quand l'évaluation engendre des échecs non justifiés
- ~~6. Quelles sont les conceptions des enseignants eu égard à l'évaluation des élèves ?~~
- ~~7. Les élèves comprennent-ils les enjeux des diverses fonctions évaluatives ?~~
8. Quand l'évaluation soutient l'apprentissage des élèves
9. En conclusion : un double changement paradigmatique pour lutter contre l'échec scolaire

1. Introduction

Introduction

Une analogie emblématique

Les deux faces de l'évaluation



Plus actuelle ?



Introduction



Contribuer à la qualité d'apprentissage



« A l'école, ce qui fait légitimement débat, ce sont **les usages pédagogiques et sociaux de l'évaluation**, entre une évaluation des élèves qui a pour fonction de **soutenir leurs apprentissages** et de les aider à progresser quand ils rencontrent une difficulté versus une évaluation qui sert à **classer les élèves** et à les hiérarchiser dans un but plus ou moins avoué de sélection et de reproduction sociales » (Crahay, Mottier Lopez & Marcoux, 2019, p. 358).



Établir des hiérarchies d'excellence



2. Des fonctions multiples, mais une visée de régulation

Des fonctions multiples, mais une visée de régulation

Fonction	Définition (Récolte d'informations permettant de ...)	Visée
Formative	Adapter les activités d' <u>enseignement</u> et d' <u>apprentissage</u> afin de <u>favoriser la progression</u> des apprenants vers les objectifs de formation	Régulation: enseignement ou apprentissage
Sommative (Certificative)	Établir un <u>bilan certifiant</u> les compétences et/ou les connaissances acquises par l'apprenant à la <u>fin d'une étape</u> d'un cycle de formation.	Régulation : atteinte des objectifs
Pronostique	Guide les <u>décisions d'admission</u> ou d'orientation de l'apprenant vers une nouvelle étape ou un nouveau cursus de formation.	Régulation : parcours de formation

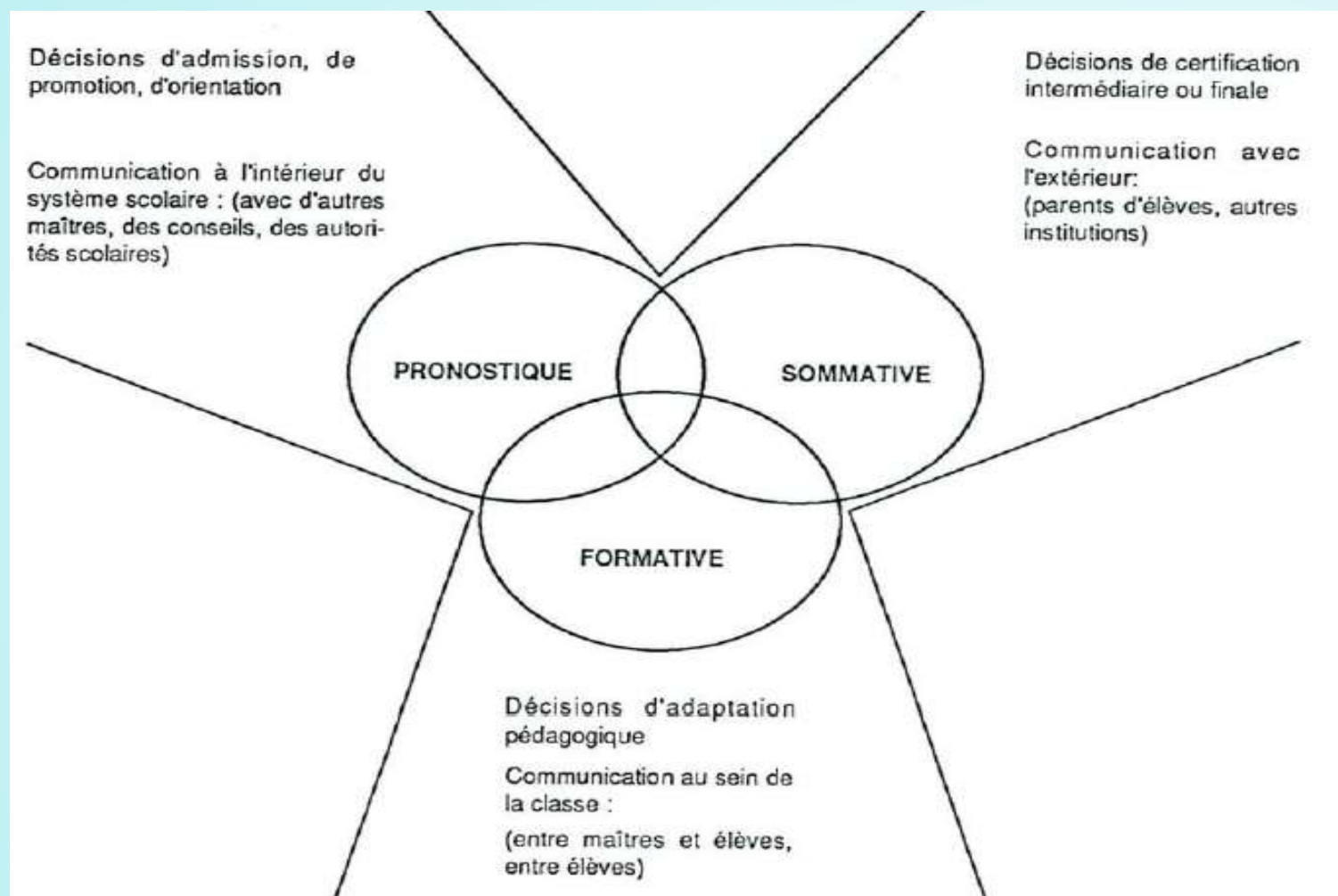
« Les modalités d'évaluation adoptées par un système de formation ont toujours une fonction de régulation, c'est-à-dire qu'elles ont pour but d'assurer l'articulation entre les caractéristiques des personnes en formation, d'une part, et les caractéristiques du système de formation, d'autre part » (Allal, 1991, p. 48).

Des fonctions multiples, mais une visée de régulation

Les procédures et outils de l'évaluation doivent être pensés de façon à permettre a décision vise et à en assure la qualité.

Modalité	Procédure
Formative	Les questions et tâches évaluatives seront ciblées sur les apprentissages en cours . Plus précisément, on concevra l'évaluation, instrumentée ou non, de manière à diagnostiquer précisément quelles sont les composantes de l'apprentissage qui posent problème à des fins de régulation.
Sommative (Certificative)	Les questions et tâches évaluatives couvrent l'éventail des objectifs qui sont considérés comme fondamentaux dans l'étape de formation concernée.
Pronostique	Les questions et les tâches évaluatives anticipent les apprentissages qui seront abordés ultérieurement ; il peut parfois être logique d'y privilégier des questions difficiles et hiérarchisant les élèves.

Des fonctions multiples, mais une visée de régulation



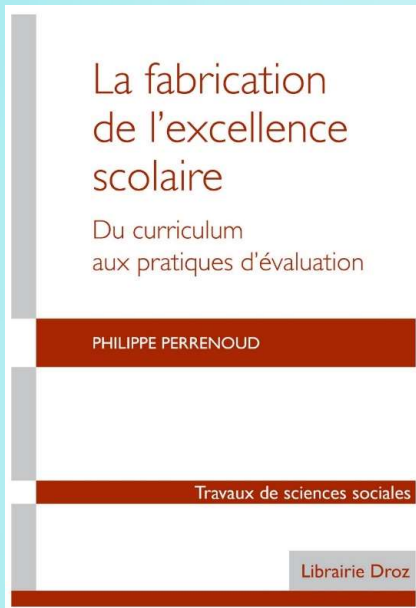
3. Confusions entre les fonctions de l'évaluation

Confusions entre les fonctions de l'évaluation

Confusion	Exemple
Formative Vs. Pronostique	Réponses à des tests formatifs utilisées pour prédire l'avenir scolaire des élèves
Sommative Vs. Pronostique	Examens pronostics portant sur des compétences issues de programmes antérieurs et non sur les prérequis Épreuves sommatives constitués de questions discriminantes et/ou portant sur des compétences non encore développées
Formative Vs. Sommative	Notation des épreuves formatives Épreuves sommatives calquées sur l'évaluation formative et/ou aboutissant sur de la remédiation

4. Évaluer dans une perspective d'excellence ou de maîtrise ?

Évaluer dans une perspective d'excellence ou de maîtrise ?



L'école a pour tradition de fabriquer des hiérarchies d'excellence.

Évaluation à référence normative

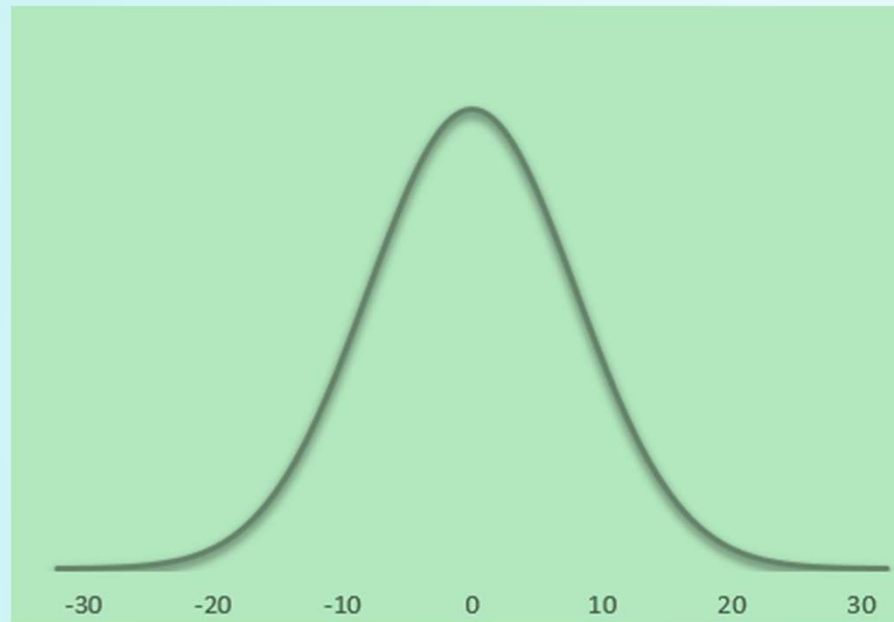
« toute performance d'élève est évaluée par comparaison avec celles d'individus placés dans la même situation afin de classer le sujet parmi ceux-ci » (de Landsheere, 1980, p. 295)

// psychométrie des tests de connaissances ou d'intelligence
→ La performance d'un individu est jugée par référence à celle d'autrui en le situant dans une distribution statistique

Hiérarchisation des élèves

Évaluer dans une perspective d'excellence ou de maîtrise ?

- 1° Questions discriminantes/épreuves obligeant les élèves à risque à se dévoiler.
- 2° Création d'une échelle artificielle tendant vers une distribution gaussienne des notes.

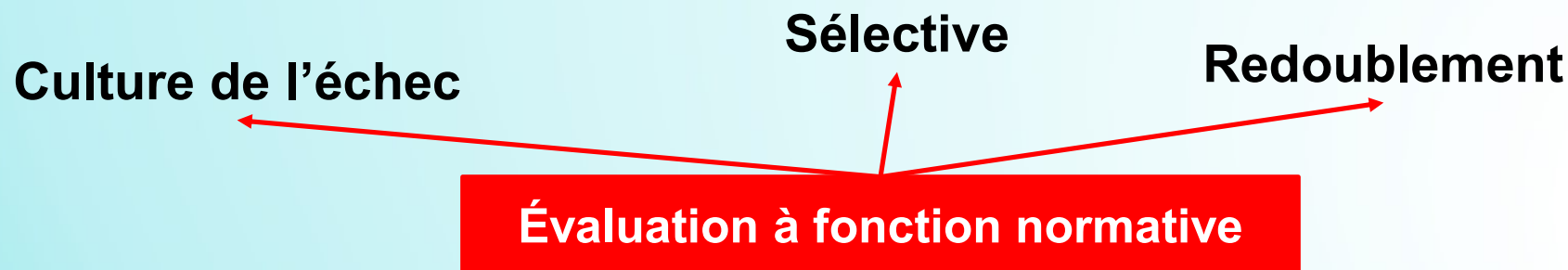


Évaluation à fonction normative

Évaluer dans une perspective d'excellence ou de maîtrise ?

- 1° Questions discriminantes/épreuves obligeant les élèves à risque à se dévoiler.
- 2° Création d'une échelle artificielle tendant vers une distribution gaussienne des notes.

« Je me dis, maintenant, en fin de 3P : est-ce que cet élève sera capable d'attaquer le programme de 4P avec toutes ses nouvelles matières ? C'est ma réponse à cette question qui va influencer ma décision. » (Enseignante interrogée par Burdevet (1994))

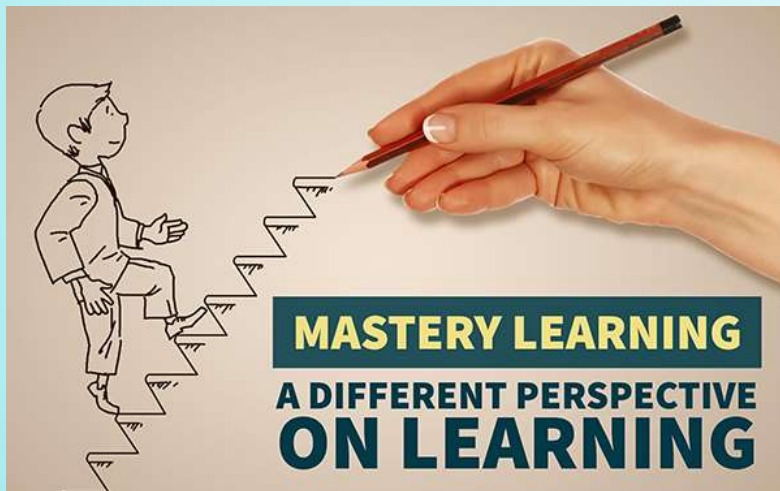


Évaluer dans une perspective d'excellence ou de maîtrise ?

Pédagogie de la maîtrise

La mission de l'enseignant est de susciter un maximum d'apprentissages chez un maximum d'élèves et donc d'amener la quasi-totalité des élèves à un même niveau élevé de réussite, et ce, malgré leurs différences initiales.

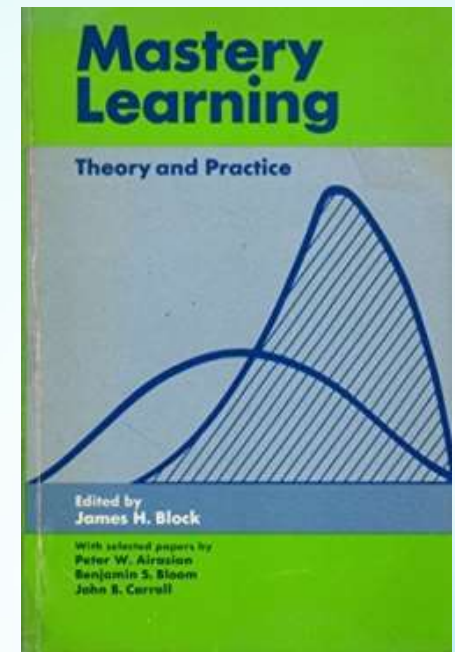
Adaptation du processus d'enseignement



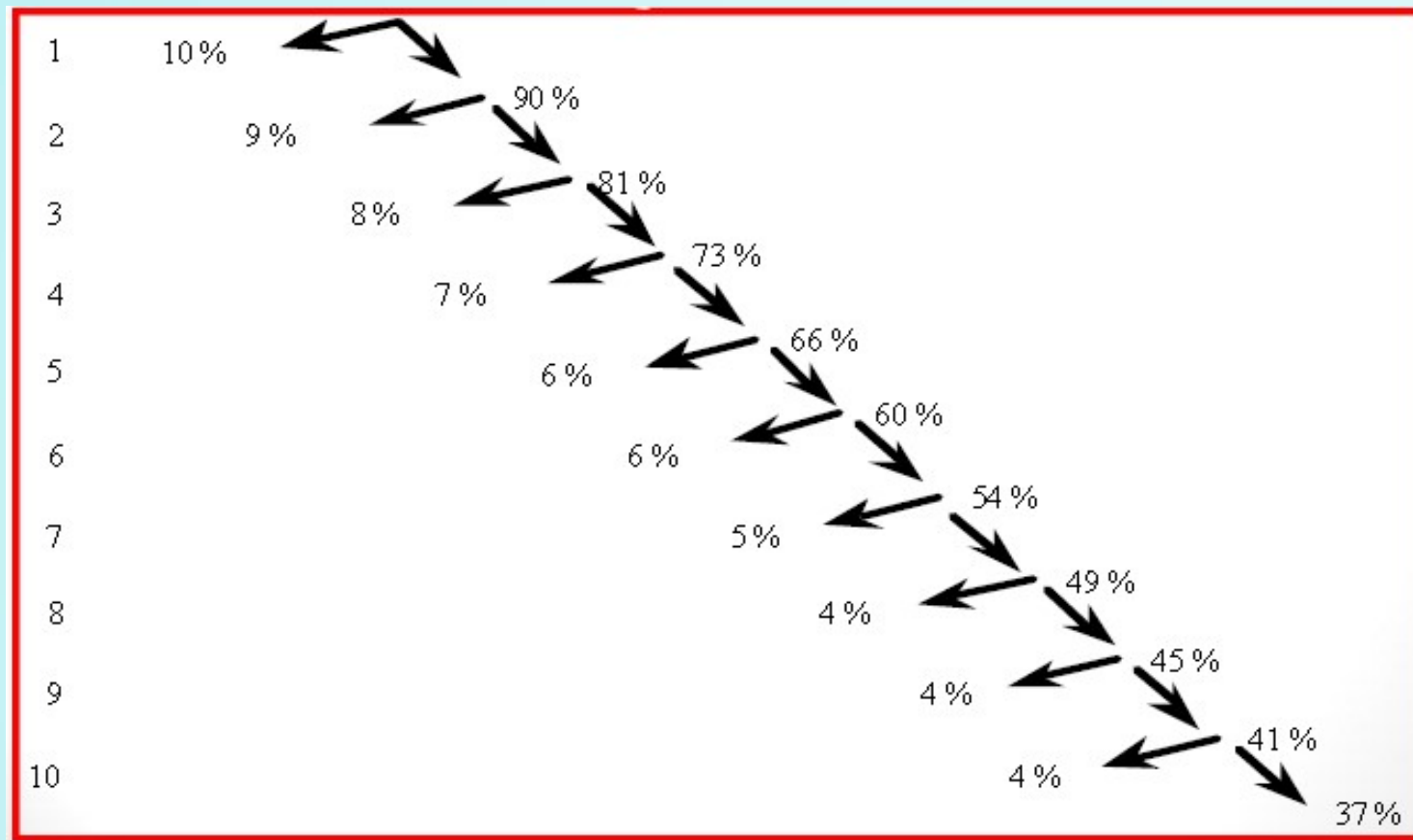
ARTICULATION

**Évaluation
sommativ à
référence critériée**
Situer l'élève par rapport
aux compétences à
maîtriser

Évaluations formatives en
cours à caractère **diagnostic**



Évaluer dans une perspective d'excellence ou de maîtrise ?



Évaluer dans une perspective d'excellence ou de maîtrise ?

Évaluation dans une perspective ...	
D'excellence	De maîtrise
Évaluation normative → Classement des individus les uns/aux autres → Distribution gaussienne des notes → Questions discriminantes	Évaluation sommative à référence critériée → Permettre de situer chaque élève par rapport aux compétences assignées comme objectifs
Évaluation pronostique → Apte à suivre ? → Sélection.	Évaluation formative → Comprendre les difficultés éprouvées par l'élève <ul style="list-style-type: none">• Diagnostic• Analyse des réponses• Interactions avec les élèves → Réorienter l'apprentissage et/ou l'enseignement

5. Quand l'évaluation engendre des échecs non justifiés

La recherche APER

A. Grisay (1984)

- Menée dans 53 écoles (n = 1503 élèves) de la FW-B (région liégeoise), elle porte sur la 5^e année de l'enseignement primaire.
- Les chercheurs recueillent dans chaque classe les notes attribuées par l'enseignant à chacun des élèves lors de l'examen de français ;
- Les chercheurs soumettent les élèves à une épreuve commune externe de français.

La recherche APER

Standardisation des notes (note Z) pour rendre comparables les résultats de l'épreuve externe et ceux l'examen de l'enseignant.

$$Z = \frac{X - M}{S}$$

- La standardisation permet de positionner chaque note ou score par rapport à la moyenne en termes d'écarts par rapport à la moyenne (ou fraction d'écart-type)
- Une note Z positive = note supérieure à la moyenne
- Une note Z négative = note inférieure à la moyenne

La recherche APER

Standardisation des notes (note Z) pour rendre comparables les résultats de l'épreuve externe et ceux l'examen de l'enseignant.

$$Z = \frac{X - M}{S}$$
$$Z = \frac{X_i - \mu_x}{\sigma_x}$$

Test de français :

$$\mu_x = 100$$

$$\sigma_x = 10$$

Sore de Thomas = 120

Score de Simon = 70

Test de français :

$$\mu_x = 100$$

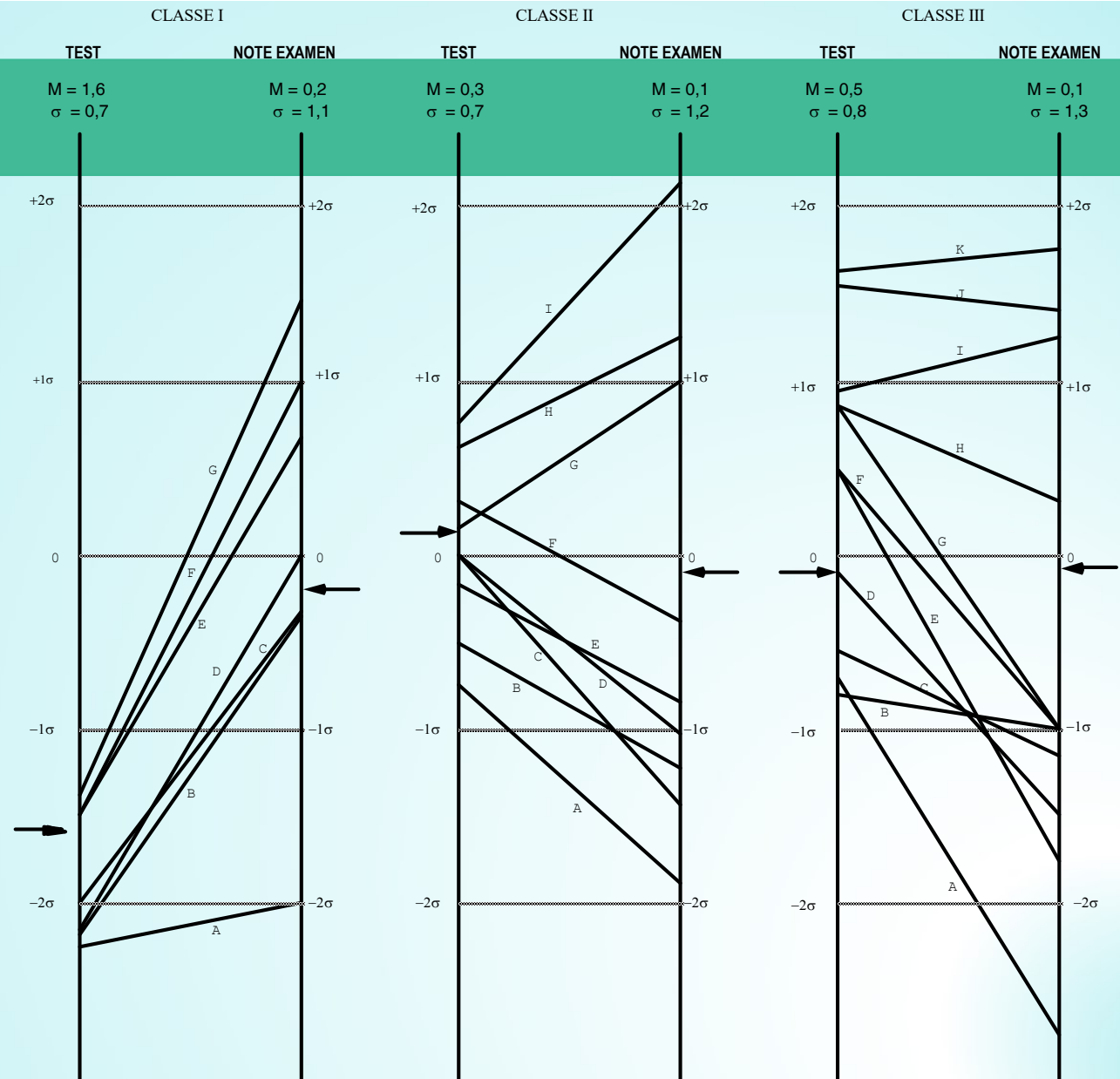
$$\sigma_x = 10$$

Note Z de Thomas = +2

Note Z de Simon = - 3

La recherche APER

Notes standardisées des
élèves de 3 classes 5^e
primaire



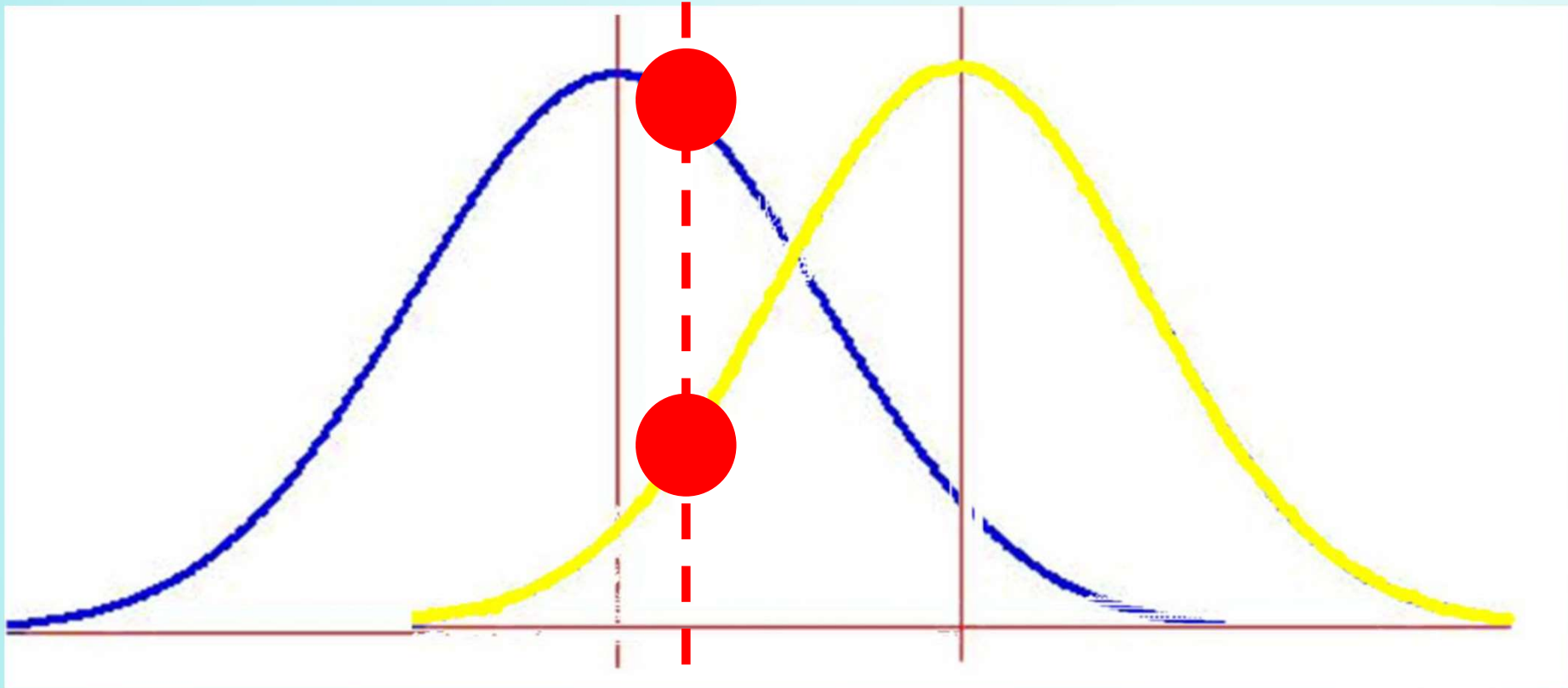
- Par rapport au « test » des chercheurs, comment qualifiez-vous la classe I, II, III ?
- Par rapport à la « note d'examen » de l'enseignant, comment qualifiez-vous les résultats de la classe I, II, III ?
- Pour chacune des trois classes, quelle est la différence entre le jugement issu des résultats d'un test externe (chercheurs) et d'un test interne (enseignant)
- Intéressez-vous à l'élève A de la classe II et de la classe III. Ces deux élèves, derniers de leur classe, vont redoubler. Si ces élèves étaient dans la classe I, pensez-vous qu'ils redoubleraient ?

La recherche APER

Expérience reproduite par Grisay, De Bal et De Landsheere (1984)

Note Z au test	Promus	Redoublants
< -3.00	4	0
[-3.00 ; -2.00 [20	17
[-2.00 ; -1.00 [130	29
[-1.00 ; - 0.50 [149	16
[- 0.50 ; 0.00 [212	10
> 0	773	4

La recherche APER



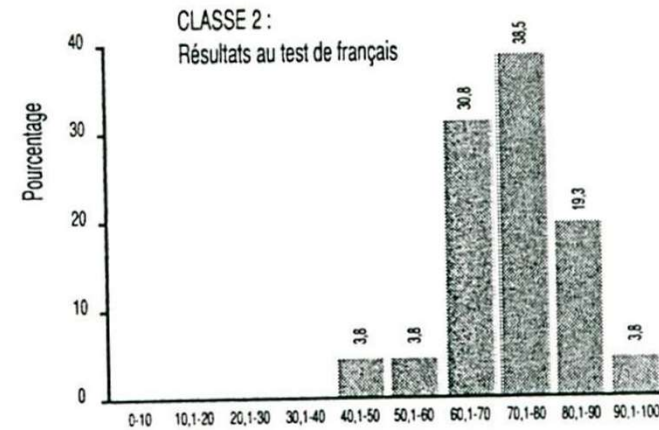
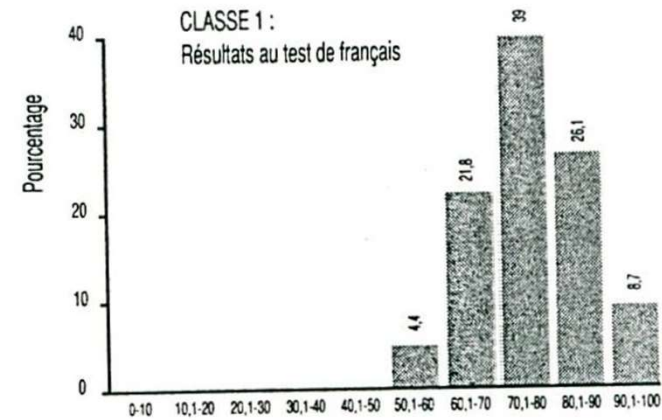
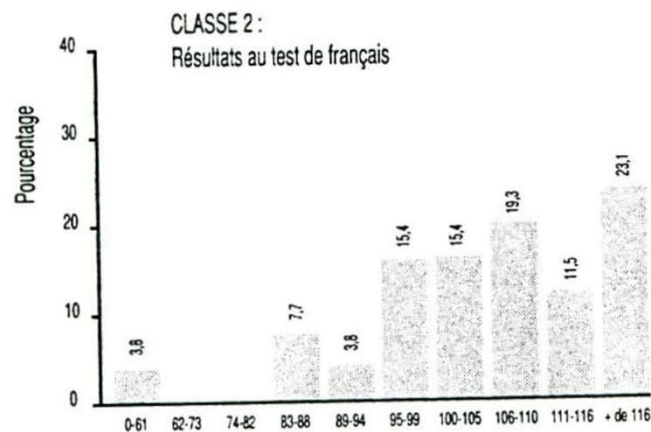
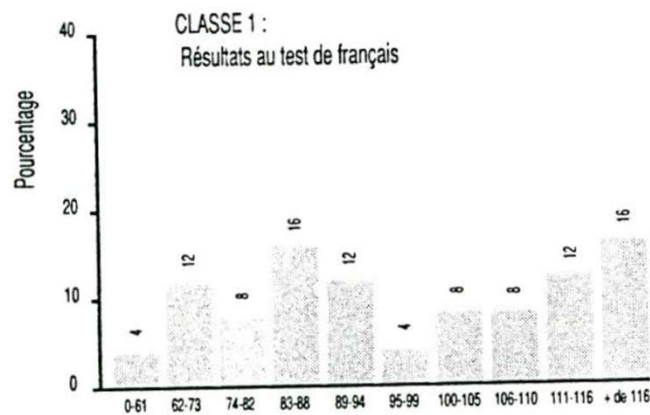
La recherche APER

Conclusions

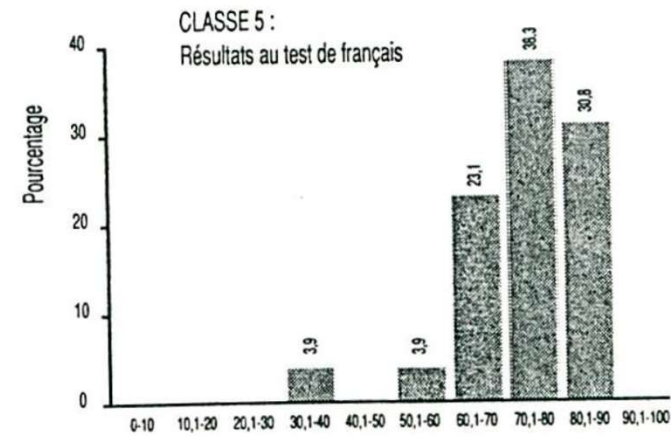
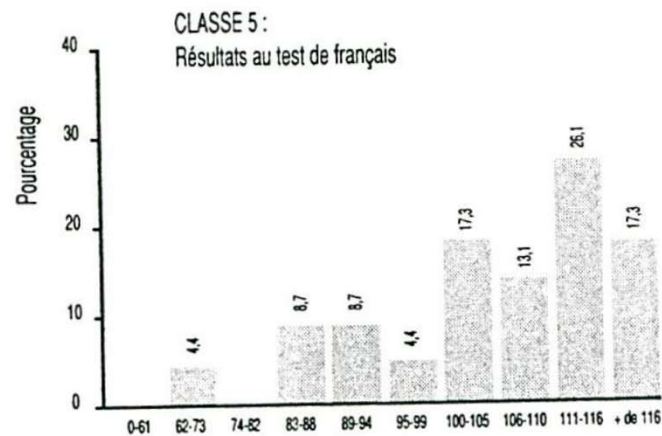
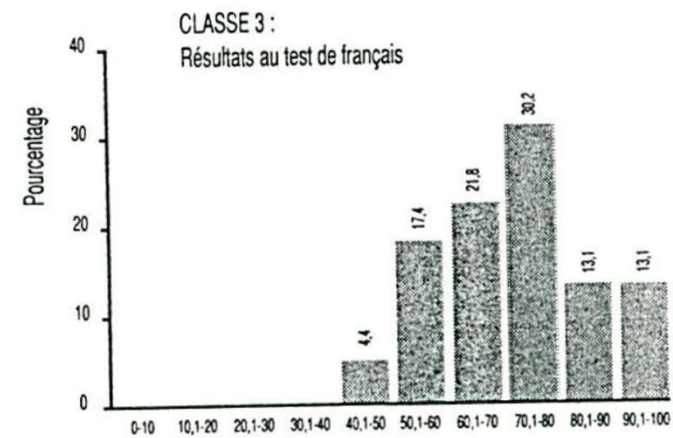
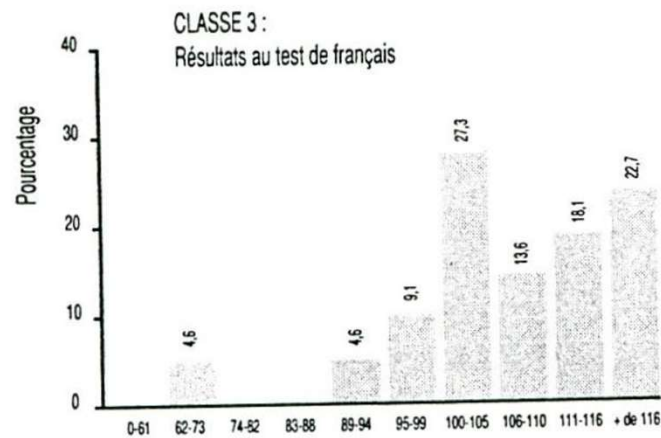
- « Lorsqu'on compare, classe par classe, les résultats exprimés en notes réduites, on constate que l'écart-type des notes à l'examen est supérieur à celui des résultats au test de français. En termes courants, ceci signifie que, dans toutes les classes, la dispersion des notes scolaires est supérieure à celle des résultats à l'épreuve externe. »
- « Tout se passe comme si l'évaluation pratiquée par les maîtres produisait une amplification de l'hétérogénéité des performances des élèves. »
- À résultats équivalents au test, un élève d'une classe va doubler et un autre appartenant à une autre classe va réussir.
- À compétence égale, un élève sera jugé fort ou faible selon la classe qu'il fréquente et donc selon la compétence des autres élèves.

Loi de Posthumus ou mirages de l'évaluation

ECOLE A

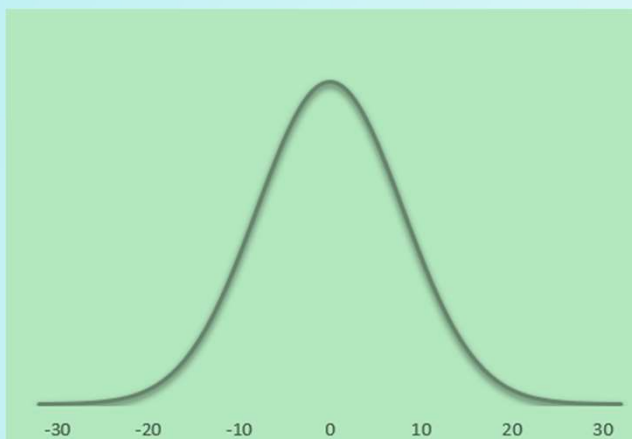


Loi de Posthumus ou mirages de l'évaluation



Loi de Posthumus ou mirages de l'évaluation

« Quelle que soit la distribution des compétences au début de l'année scolaire, la distribution des notes en fin d'année épouse *grosso modo* une forme gaussienne (allure de la courbe normale) »



Loi de Posthumus

« un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver d'année en année, approximativement, la même distribution gaussienne) des notes » (de Landsheere, 1980, p. 224)

Loi de Posthumus ou mirages de l'évaluation

« Les trois exemples illustrent un même mécanisme d'adaptation de la note scolaire au niveau de performance de la classe. Ils mettent bien en évidence le caractère tout relatif d'une notion comme celle d'hétérogénéité, sur laquelle se fonde, en principe, le redoublement. Une classe tout à fait homogène peut parfaitement paraître disparate à l'enseignant, puisque l'examen et l'échelle d'évaluation qu'il utilise sont adaptés au niveau des élèves : il ne va pas enregistrer, par exemple dans le cas de la classe 1, la faiblesse commune qui fonde l'homogénéité du groupe ; il sera par contre sensible à des variations entre élèves plus fines que celles enregistrées par le test » (Grisay, 1984, p. 35).

Loi de Posthumus ou mirages de l'évaluation

Répartition de la variance entre classes et entre élèves dans les classes (1503 élèves et 53 classes) (Grisay *et al.*, 1984)

	Score au test	Score à l'examen
Variations entre les élèves	57,9 %	82,8 %
Variations entre classes	42,1 %	17,2 %

Certes, lorsqu'on dispose d'une mesure fiable d'un paramètre et d'un échantillon aléatoire et simple, on obtient une distribution dite normale (courbe de Gauss)

Or, les classes sont rarement composées selon un échantillonnage aléatoire et simple

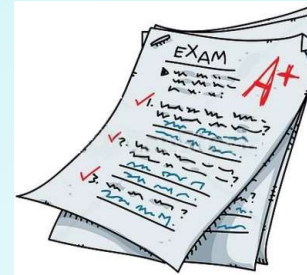
- la composition des classes n'est pas soumise à la loi des grands nombres
- On observe plutôt un effet d'agrégation selon l'origine sociale et selon les compétences académiques.
- Il n'y a aucune raison qu'on retrouve une distribution gaussienne des notes au sein des classes.

Bonus : expérience de Gjorgjevski (cité par Rot et Butas, 1959)

Étape 1



5 correcteurs

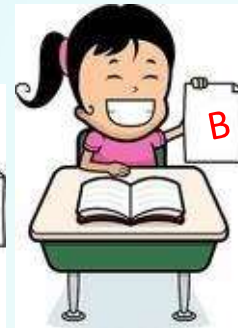
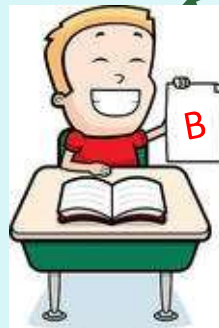


100 épreuves

Étape 2



4 autres correcteurs

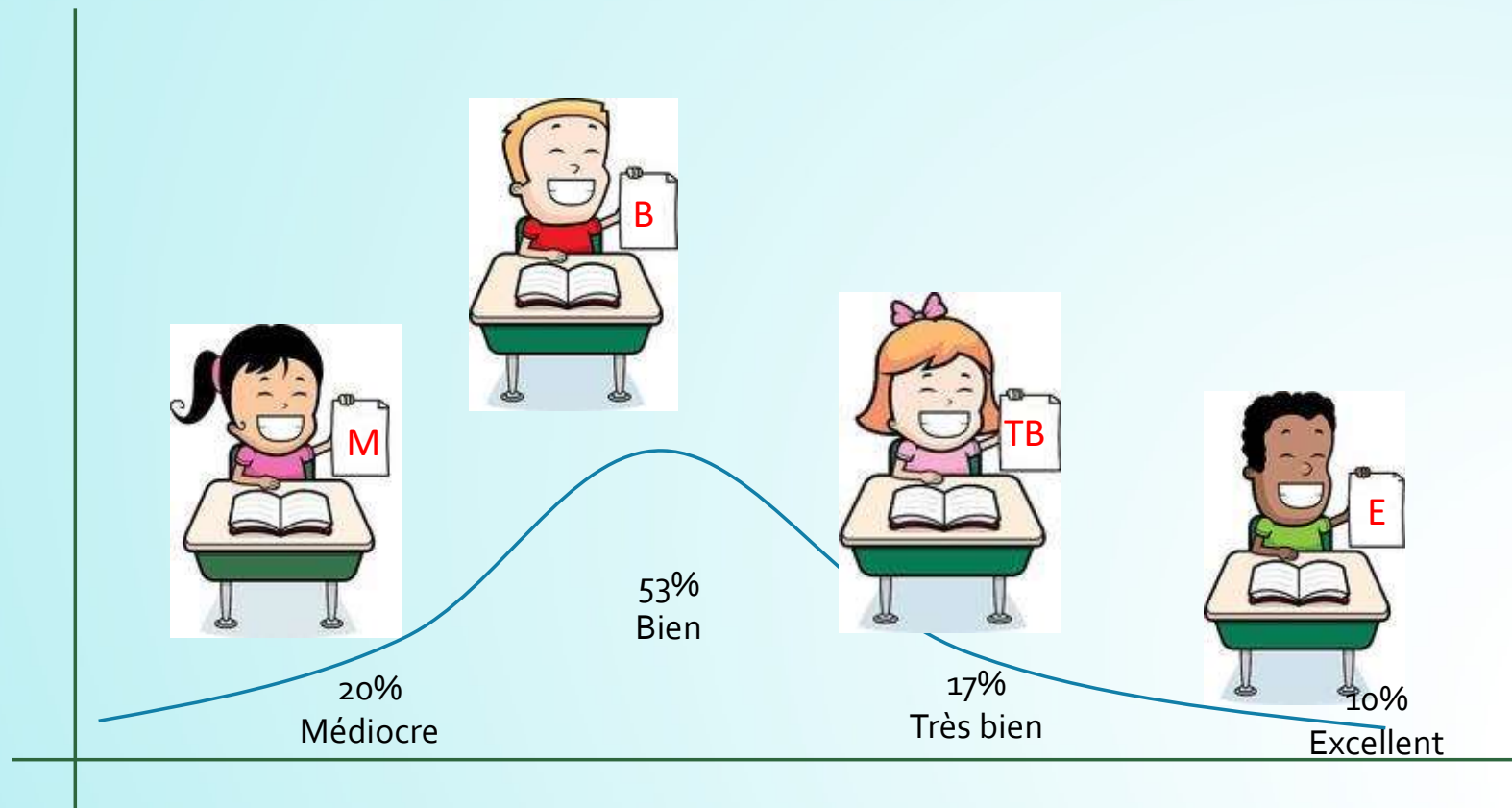


15 copies notées « Bien »

Bonus : expérience de Gjorgjevski (cité par Rot et Butas, 1959)

Résultats

Les nouveaux correcteurs ont spontanément adoptés des exigences nouvelles



Diversité des exigences des enseignants

APER : étude des questions d'examen conçues par les enseignants

- Le degré d'exigence attendu des élèves d'un même niveau scolaire varie de façon considérable d'une classe à l'autre.
- Pour passer de la 3^e à la 4^e primaire, il faut, dans certaines écoles, maîtriser les objectifs relevant de la 4^e (voire de la 5^e) primaire alors que, dans d'autres, la maîtrise des objectifs de fin de 2^e est suffisante.

Detheux & Kellens (1992) : étude d'un corpus de 2 000 questions de français

- Diversité considérable dans la difficulté des questions supposées évaluer le même objectif.

Diversité des exigences des enseignants

Répartition des points établie par 16 professeurs de français de première secondaire entre les différentes sections que doit comporter un examen de fin d'année (Max = 60) (Detheux & Kellens, 1992)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Grammaire	25	30	28	22	21	30	29	38	23	23	39	26	30	14	27	29
Orthographe	10	15	13	17	22	2	19	11	13	16	13	9	17	8	13	11
Vocabulaire	13	1	0	1	9	7	1	0	0	7	2	2	3	6	0	0
Stylistique	3	3	0	0	0	9	1	0	6	2	1	0	1	0	0	0
Comp. textes	4	7	13	10	8	5	6	9	15	0	0	13	3	13	8	17
Prod. Ecrite	5	4	6	10	0	7	4	2	3	12	5	10	6	19	12	3
Total	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60

- Comparez les enseignants 13 et 14.
- Valorisent-ils les mêmes aspects du français ?
- Chez quel enseignant vaut-il mieux être si on a des difficultés en grammaire ?
- Quelle est la place de la compréhension en lecture et de la production écrite dans ce panel de 16 enseignants de 1^{re} secondaire ?

Les élèves en difficulté sont-ils victimes d'un effet de stéréotypie ?

Effet de stéréotypie = tendance à attribuer la même note aux élèves malgré des variations de performance significatives de leur part.

Corrélations entre les notes de français aux différentes périodes de l'année scolaire pour l'ensemble des classes (Bounafaa, 1992)

	Décembre	Noël	Février	Mars	Mai	Juin	Fin
Octobre	0.62	0.74	0.65	0.64	0.59	0.94	0.80
Décembre		0.76	0.67	0.71	0.69	0.68	0.66
Noël			0.76	0.77	0.77	0.80	0.84
Février				0.80	0.82	0.91	0.81
Mars					0.88	0.91	0.81
Mai						0.93	0.80
Juin							0.87

Quand l'évaluation soutient l'apprentissage des élèves

Les élargissements conceptuels de l'évaluation formative

Conception initiale de Bloom versus conception élargie de l'évaluation formative dans la littérature scientifique francophone (Allal & Mottier Lopez (2005))

Conception initiale de Bloom	Conception élargie
Insertion de l'EF <u>après</u> la phase d'enseignement	Intégration de l'EF durant <u>tout</u> l'apprentissage
Utilisation de <u>tests</u> formatifs	Utilisation de <u>divers moyens</u> de recueil d'informations
Feed-Back + correction → <u>remédiation</u>	Feed-back + adaptation de l'enseignement → <u>régulation</u>
Gestion de l'EF par l' <u>enseignant</u>	Participation active des <u>élèves</u> à l'EF
Maîtrise des objectifs par <u>tous les élèves</u>	<u>Différenciation</u> de l'enseignement et des objectifs
<u>Remédiation bénéfique</u> aux élèves qui ont été évalués	<u>Régulation à 2 niveaux</u> : pour les élèves évalués et pour les futurs élèves

Impliquer les élèves dans des démarches d'auto-évaluation

Étude expérimentale de Nunziati (1990) :

- Impact positif quand les élèves sont en mesure de s'approprier les critères d'évaluation des enseignants, d'autogérer leurs erreurs et de maîtriser des outils d'anticipation et de planification de l'évaluation en classe.

Évaluation formative	Évaluation formatrice
Peut être à référence critériée	
Adapter la phase d'enseignement/d'apprentissage	Impliquer l'élève dans l'ensemble du processus d'évaluation formative Apprendre à s'autoévaluer correctement
Gérée exclusivement par l'enseignant	Gérée exclusivement par l'élève
Coresponsabilité entre les enseignants et les élèves dans des modalités dynamiques et complémentaires	

L'évaluation informelle

Évaluation formative formelle

Conçue pour produire des traces et indices sur l'apprentissage de l'élève
→ *épreuves papier-crayon débouchant sur un feed-back critérié explicite*

Recueillir

Des informations sur les élèves
dans un temps planifié

Interpréter (analyser)

Les informations recueillies
(souvent hors de la présence des
élèves)

Agir

en planifiant une action visant à
soutenir l'élève dans l'atteinte des
buts d'apprentissage

Évaluation formative informelle

Traces et indices générées pendant l'activité quotidienne de classe
→ Pleinement intégrées aux processus d'enseignement et d'apprentissage en contexte de classe

Solliciter

Des réponses verbales des élèves
afin de récolter de l'information

Reconnaître

Les réponses des élèves au
regard des concepts enseignés

Utiliser

Immédiatement l'information dans
le cours des activités continues de
la classe

Une évaluation formative foncièrement contextualisée

« l'environnement social de la classe est de plus en plus considéré comme un facteur important et une conséquence (*outcome*) de comment les enseignants mettent en œuvre différentes pratiques évaluatives. L'environnement social s'applique particulièrement aux interventions qui visent à promouvoir des productions d'élèves (orales ou écrites) en tant que traces qui montrent leurs raisonnements et compréhension, et qui sont susceptibles d'éclairer et de renseigner l'enseignant » (Young & Kim, 2010, p. 16)

En conclusion : un double changement paradigmatique pour lutter contre l'échec scolaire

En conclusion

La lutte contre l'échec scolaire appelle un double changement :

1. Abandon de l'évaluation normative ➔ articulation de l'évaluation formative et sommative à référence critériée.
2. Intégration de la notion de régulation au cœur de l'évaluation formative.

Vrai/Faux ?

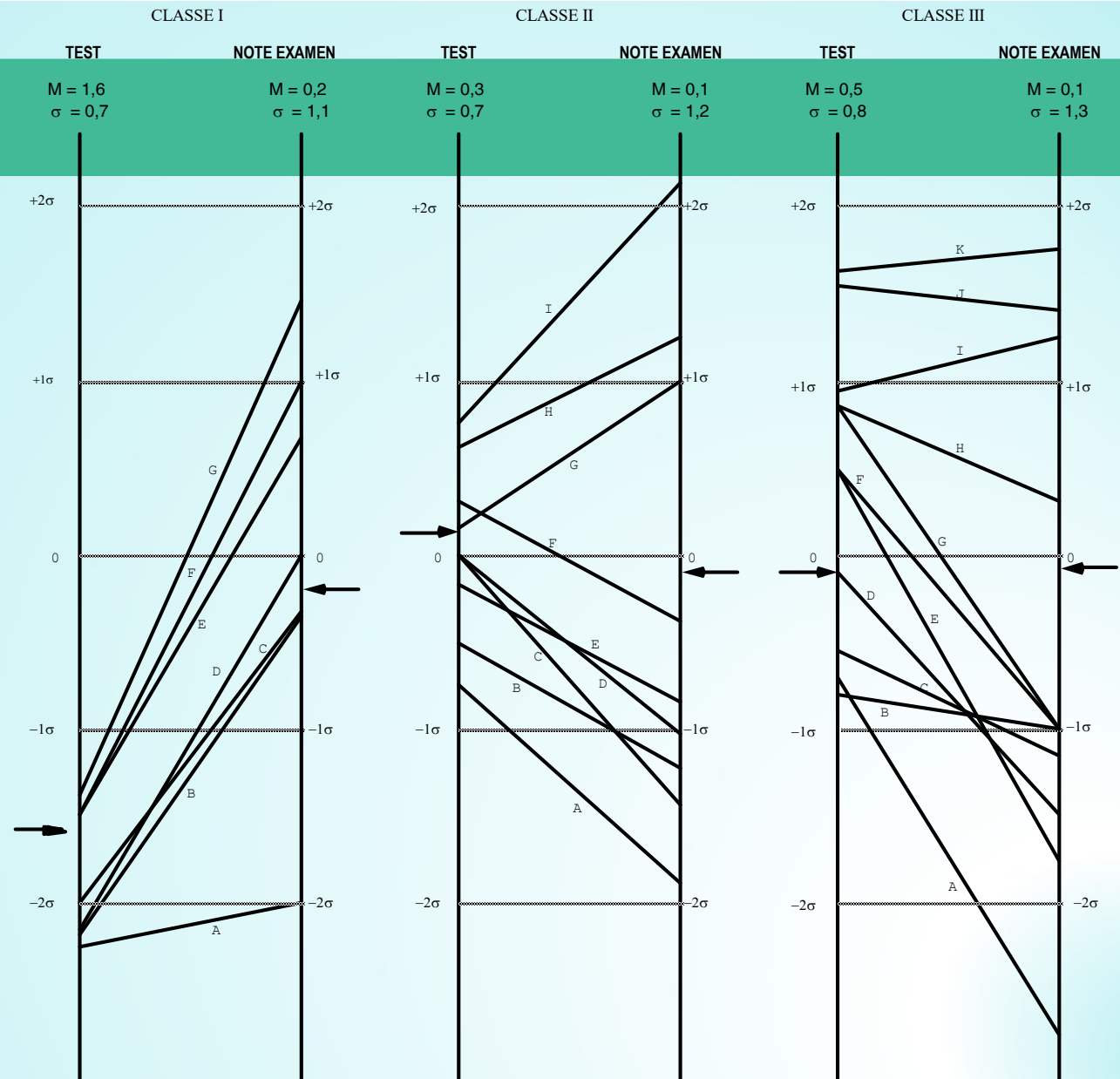
Vrai/Faux ?

La classe 1 est nettement plus faible que la classe 2.

Vrai/Faux ?

Dans les trois classes, la dispersion des notes à l'examen de l'instituteur est inférieure à celle des notes à l'épreuves externes.

Vrai/Faux ?



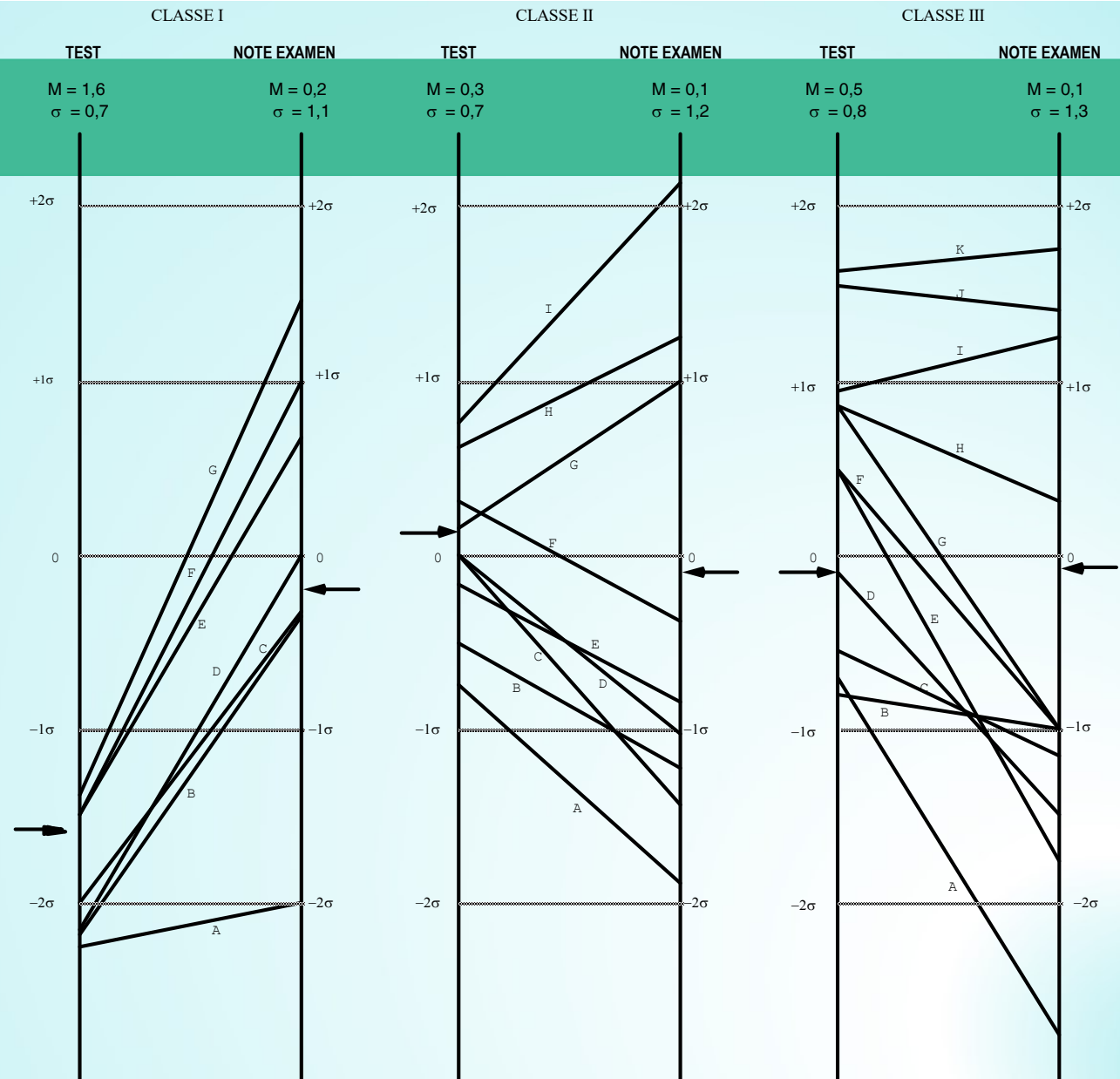
Vrai/Faux ?

La classe 1 est nettement plus faible que la classe 2.

Vrai

Dans les trois classes, la dispersion des notes à l'examen de l'instituteur est inférieure à celle des notes à l'épreuves externes.

Vrai/Faux ?



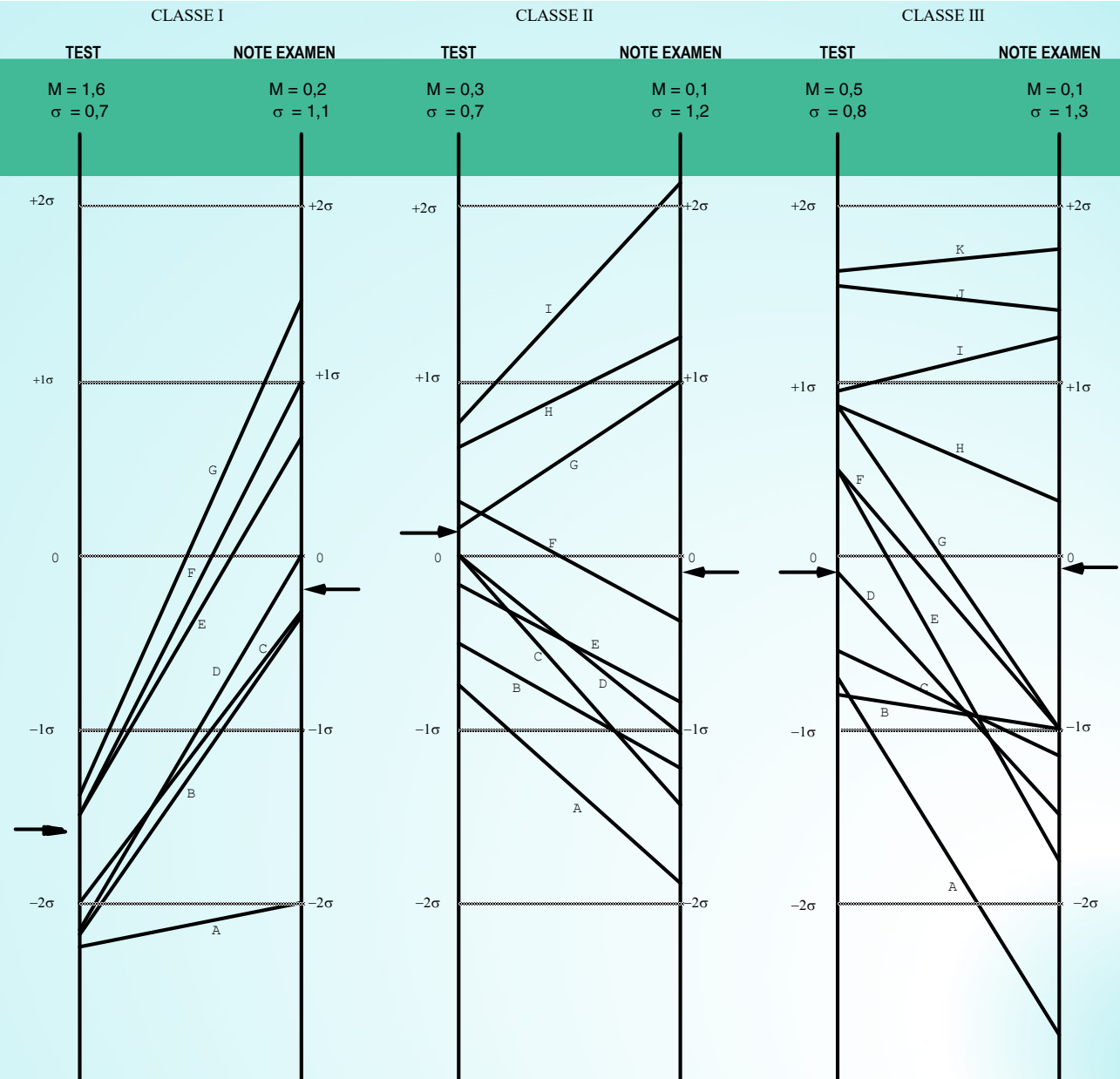
Vrai/Faux ?

La classe 1 est nettement plus faible que la classe 2.

Vrai

Dans les trois classes, la dispersion des notes à l'examen de l'instituteur est inférieure à celle des notes à l'épreuves externes.

Faux



Vrai/Faux ?

L'effet Posthumus a pour conséquence que la note qu'un élève obtient dans une classe est tributaire du niveau de cette classe.

Vrai/Faux ?

L'effet Posthumus a pour conséquence que la distribution des notes des élèves d'une classe suit approximativement une courbe en « j ».

Vrai/Faux ?

Vrai/Faux ?

L'effet Posthumus a pour conséquence que la note qu'un élève obtient dans une classe est tributaire du niveau de cette classe.

Vrai

L'effet Posthumus a pour conséquence que la distribution des notes des élèves d'une classe suit approximativement une courbe en « j ».

Faux

Vrai/Faux ?

En FW-B, les élèves qui redoublent sont relativement homogènes du point de vue de leur compétence scolaire.

Vrai/Faux ?

De manière générale, les enseignants de FW-B s'accordent sur leurs exigences, notamment pour ce qui est de la pondération des différentes matières.

Vrai/Faux ?

Vrai/Faux ?

En FW-B, les élèves qui redoublent sont relativement homogènes du point de vue de leur compétence scolaire.

Faux

De manière générale, les enseignants de FW-B s'accordent sur leurs exigences, notamment pour ce qui est de la pondération des différentes matières.

Vrai/Faux ?

Vrai/Faux ?

En FW-B, les élèves qui redoublent sont relativement homogènes du point de vue de leur compétence scolaire.

Faux

De manière générale, les enseignants de FW-B s'accordent sur leurs exigences, notamment pour ce qui est de la pondération des différentes matières.

Faux

Vrai/Faux ?

Une évaluation normative permet d'obtenir, au sein de sa classe, une distribution gaussienne des notes.

Vrai/Faux ?

Avec l'évaluation normative, les enseignants adaptent le niveau d'exigence des questions aux capacités de la majorité supérieure des élèves.

Vrai/Faux ?

Vrai/Faux ?

Une évaluation normative permet d'obtenir, au sein de sa classe, une distribution gaussienne des notes.

Vrai

Avec l'évaluation normative, les enseignants adaptent le niveau d'exigence des questions aux capacités de la majorité supérieure des élèves.

Vrai/Faux ?

Vrai/Faux ?

Une évaluation normative permet d'obtenir, au sein de sa classe, une distribution gaussienne des notes.

Vrai

Avec l'évaluation normative, les enseignants adaptent le niveau d'exigence des questions aux capacités de la majorité supérieure des élèves.

Vrai

Vrai/Faux ?

En utilisant l'évaluation sommative à référence critériée, l'école fabrique des hiérarchies d'excellence.

Vrai/Faux ?

L'évaluation formative a pour fonction d'adapter les activités d'enseignement et d'apprentissage afin de favoriser la progression des apprenants vers les objectifs de formation

Vrai/Faux ?