



ΚΕ-ΓΛΩ-21
Αξιολόγηση δεξιοτήτων
επικοινωνίας
στις ξένες γλώσσες

ΚΕ-GLO-21
Évaluation des compétences
de communication
en langue étrangère

Ενότητα 4
Abord d'activités de recherche en évaluation

Olivier DELHAYE
Τμήμα Γαλλικής γλώσσας και φιλολογίας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



Άδειες Χρήσης

**Το παρόν εκπαιδευτικό υλικό υπόκειται
σε άδειες χρήσης Creative Commons.**

**Για εκπαιδευτικό υλικό,
όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης,
η άδεια χρήσης αναφέρεται ρητώς.**



Χρηματοδότηση

Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.

Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.

Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





Évaluation des compétences de communication en langue étrangère

Module 4

Abord d'activités de recherche en évaluation

4 Abord d'activités de recherche en évaluation

- 4.1 Éléments d'analyse d'items
- 4.2 Analyse de la distribution des notes
- 4.3 Établissement de corrélations



4.1 Éléments d'analyse d'items

- 4.1.1 Score, note, certification
- 4.1.2 Indice de facilité
- 4.1.3 Indice de discrimination
- 4.1.4 Établissement de l'efficacité
- 4.1.5 Exemple d'application



4.1.1 Score, note, certification

Les étapes du processus de l'évaluation certificative sont les suivantes (Figure 80). Nous prendrons l'exemple d'épreuves des examens du ΚΠγ.

- Intrans : scores (liste des bonnes et des mauvaises réponses de chaque testé à chaque item / des appréciations attribuées à chaque performance et pour chaque critère)
- Traitement des scores : résultats
- Extrans : Jugement/décision

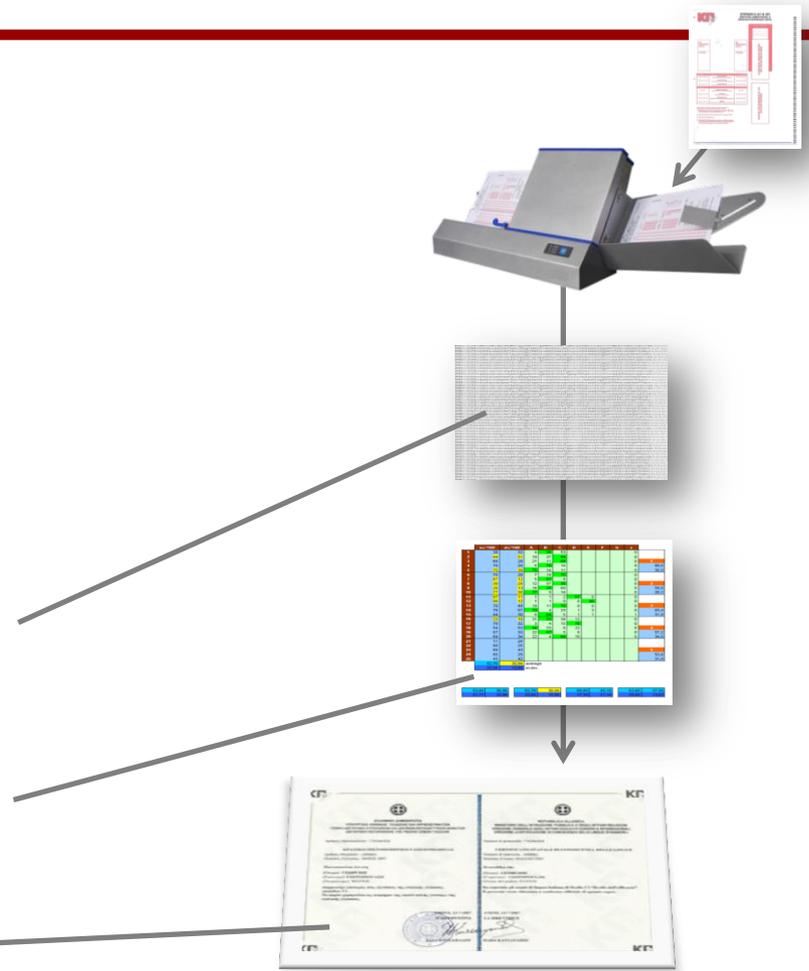


Figure 80 – Processus de l'évaluation certificative (ΚΠγ)



4.1.1 Score, note, certification

Résultats : Interprétation possible (notes) mais encore difficile

A1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A/A	ΕΠΙΠΕΔΟ	ΓΛΩΣΣΑ	ΕΝΟΤΗΤΑ	Α ΒΑΘ/ΤΗΣ	Β ΒΑΘ/ΤΗΣ	ΓΕΝ. ΣΥΝΟΛΟ	ΔΡΑΣΤ 1(A)	ΔΡΑΣΤ 2(A)	ΣΥΝΟΛΟ(A)	Α ΒΑΘ/ΤΗΣ	ΔΡΑΣΤ 1(B)	ΔΡΑΣΤ 2(B)	ΣΥΝΟΛΟ(B)	Β ΒΑΘ/ΤΗΣ	
1	1	3	2	2	17	18	17,5	9	8	17	3901	9	18	18	3905
2	2	3	2	2	26	27	26,5	13	13	26	3901	14	13	27	3905
3	3	4	2	2	11	13	24	7	4	11	3916	8	5	13	3904
4	4	2	1	2	14	14	14	8	6	14	5081	8	6	14	5015
5	5	2	1	2	18	18	18	10	8	18	5081	9	9	18	5015
6	6	2	1	2	13	12	12,5	7	6	13	5081	7	5	12	5015
7	7	2	1	2	21	14	17,5	11	10	21	5081	8	6	14	5015
8	8	2	1	2	2	2	2	1	1	2	5022	1	1	2	5078
9	9	2	1	2	22	17	19,5	11	11	22	5022	9	8	17	5078
10	10	2	1	2	23	21	22	11	12	23	5022	12	9	21	5078
11	11	2	1	2	16	18	17	8	8	16	5081	9	9	18	5015
12	12	2	1	2	17	20	18,5	9	8	17	5022	12	8	20	5078
13	13	2	1	2	27	24	25,5	13	14	27	5022	13	11	24	5078
14	14	2	1	2	21	16	18,5	11	10	21	5081	9	7	16	5015
15	15	2	1	2	13	15	14	6	7	13	5022	8	7	15	5078
16	16	2	1	2	10	15	12,5	5	5	10	5081	8	7	15	5015
17	17	1	1	2	6	11	8,5	3	3	6	5073	6	5	11	5029
18	18	1	1	2	11	13	12	6	5	11	5073	6	7	13	5029
19	19	1	1	2	19	20	19,5	10	9	19	5073	9	11	20	5029
20	20	1	1	2	10	11	10,5	5	5	10	5073	5	6	11	5029
21	21	1	1	2	20	16	18	12	8	20	5073	8	8	16	5029
22	22	1	1	2	13	15	14	6	7	13	5073	7	8	15	5029

Figure 82 – Notes (Ibid.)



4.1.1 Score, note, certification

Traitement des résultats

	e.l.*100	d.l.*100	A	B	C	D	E	F	G	x		
1	38	32	9	38	53					0		
2	44	51	19	37	44					0		
3	68	29	21	11	68					0	1	
4	78	24	6	78	16					0	60,6	
5	75	39	75	14	11					0	35,0	
6		20	7	19	75					0		
7	87	12	5	87	8					0		
8	30	26	12	57	30					0	2	
9	26	13	14	26	60					0	59,0	
10	77	30	77	9	14					0	20,2	
11	97	11	1	1	2	97	0			0		
12	96	12	1	1	0	0	96			0		
13	70	49	19	11	70	0	0			0	3	
14	70	57	70	4	23	1	2			0	83,4	
15	84	30	8	84	5	1	1			1	31,8	
16	23	18	31	23	34	12				0		
17	78	32	3	4	15	78				0		
18	54	53	54	23	0	22				0	4	
19	67	33	22	67	5	6				0	57,2	
20	64	34	22	4	64	10				0	34,0	
21	11	24										
22	60	25										
23	69	43									5	
24	85	25									53,6	
25	43	42									31,8	
	62,76	30,56	average									
	22,88	12,90	st.dev.									

53,08	36,96	62,76	30,56	68,80	40,16	62,40	37,56
21,17	10,90	22,88	12,90	17,39	21,19	25,68	13,62

Figure 82 – Notes (Ibid.)



4.1.1 Score, note, certification

	e.i.*100	d.i.*100	A	B	C	D	E	F	G	x	
1	38	32	9	38	53					0	
2	44	51	19	37	44					0	
3	68	29	21	11	68					0	1
4	78	24	6	78	16					0	60,6
5	75	39	75	14	11					0	35,0
6	75	20	7	19	75					0	
7	87	12	5	87	8					0	
8	30	26	12	57	30					0	2
9	26	13	14	26	60					0	59,0
10	77	30	77	9	14					0	20,2
11	97	11	1	1	2	97	0			0	
12	96	12	1	1	0	0	96			0	
13	70	49	19	11	70	0	0			0	3
14	70	57	70	4	23	1	2			0	83,4
15	84	30	8	84	5	1	1			1	31,8
16	23	18	31	23	34	12				0	
17	78	32	3	4	15	78				0	
18	54	53	54	23	0	22				0	4
19	67	33	22	67	5	6				0	57,2
20	64	34	22	4	64	10				0	34,0
21	11	24									
22	60	25									
23	69	43									5
24	85	25									53,6
25	43	42									31,8
	62,76	30,56	average								
	22,88	12,90	st.dev.								

53,08	36,96	62,76	30,56	68,80	40,16	62,40	37,56
21,17	10,90	22,88	12,90	17,39	21,19	25,68	13,62

Indice de facilité :
pour chaque item,
calcul de la
moyenne des scores

Indice de discrimination :
pour chaque item,
calcul de la
différence entre la
moyenne des scores
obtenus par les
meilleurs testés et la
moyenne des scores
obtenus par les
moins bons testés)

Figure 84 – Organisation des résultats (Ibid.)



4.1.2 Indice de facilité

	e.i.*100	d.i.*100
1	38	32
2	44	51
3	68	29
4	78	24
5	75	39
6	75	20
7	87	12
8	30	26
9	26	13
10	77	30
11	97	11
12	96	12
13	70	49
14	70	57
15	84	30
16	23	18
17	78	32
18	54	53
19	67	33
20	64	34
21	11	24
22	60	25
23	69	43
24	85	25
25	43	42

Figure 85 –
Indices de facilité (Ibid.)

L'indice de facilité de l'item est égal à la somme des réponses correctes divisée par le nombre des examinés.

Un bon degré de facilité tourne aux alentours de .5 ; les taux inférieurs à .33 ou supérieurs à .66 peuvent indiquer un dysfonctionnement de l'activité ou simplement une trop grande simplicité.



4.1.3 Indice de discrimination

	e.i.*100	d.i.*100
1	38	32
2	44	51
3	68	29
4	78	24
5	75	39
6	75	20
7	87	12
8	30	26
9	26	13
10	77	30
11	97	11
12	96	12
13	70	49
14	70	57
15	84	30
16	23	18
17	78	32
18	54	53
19	67	33
20	64	34
21	11	24
22	60	25
23	69	43
24	85	25
25	43	42

Figure 86 –
Indices de discrimination (Ibid.)

Calcul de l'**indice de discrimination** : on classe les examinés suivant leur score total, du meilleur au moins bon. Imaginons qu'ils sont 1000.

Pour calculer l'indice de discrimination d'un item, on calcule l'indice de facilité moyen de cet item pour les 300 meilleurs examinés (A). Puis on calcule l'indice de facilité moyen pour les 300 moins bons examinés (B). On fait $A - B$ et on obtient ainsi l'indice de discrimination.

L'indice *zéro* signifie que pour un item X, la moyenne obtenue par les bons candidats est identique à la moyenne obtenue par les mauvais candidats : l'item n'est pas discriminant, il n'a pas permis de distinguer les bons candidats des mauvais.

L'indice 1 signifie qu'absolument tous les bons candidats ont répondu correctement et qu'absolument tous les moins bons candidats ont répondu faux : l'item est incroyablement discriminant, il a permis de distinguer les meilleurs des moins bons.



4.1.4 Établissement de l'efficacité

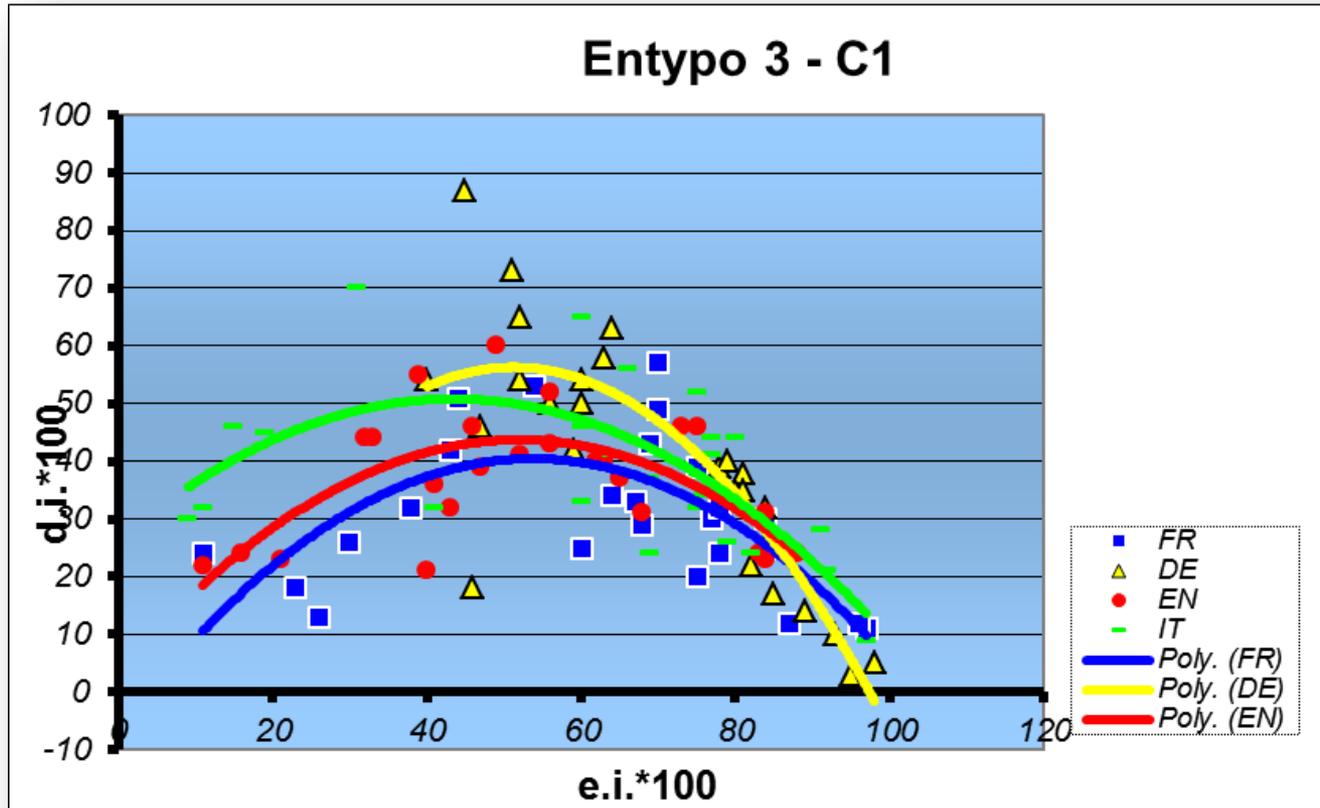


Figure 87 – Calcul de l'efficacité d'un item (Ibid.)



4.1.4 Établissement de l'efficacité

Les indices de facilité de chaque item ont été portés en abscisse (horizontalement). Les indices de discrimination en ordonnée (verticalement).

Les items qui se trouvent plus haut sont plus discriminants.

Les items qui ne se trouvent ni trop à gauche, ni trop à droite ont un degré de facilité satisfaisant.

Les items qui se trouvent dans le rectangle blanc sont problématiques : bien qu'ils ne soient ni trop faciles, ni trop difficiles, ils n'ont pas permis de distinguer les meilleurs des moins bons. Ils doivent être inspectés et , éventuellement éliminés.

Les lignes sont des courbes de tendance. Une ligne plus haute signifie que les items sont plus efficaces (facilité ok, discrimination au top).

Chaque couleur représente une langue.



4.1.5 Exemple d'application

Analyse des résultats à l'épreuve 3 du niveau C1 de la session de mai 2008 (Ministère grec de l'éducation nationale, 2008. Examens pour l'obtention du Certificat d'État de connaissance de la langue française. Niveau C1. Épreuve de compréhension de l'oral).



4.1.5 Exemple d'application

	e.i.*100	d.i.*100	A	B	C	D	E	F	G	x	
1	38	32	9	38	53					0	
2	44	51	19	37	44					0	
3	68	29	21	11	68					0	1
4	78	24	6	78	16					0	60,6
5	75	39	75	14	11					0	35,0
6	75	20	7	19	75					0	
7	87	12	5	87	8					0	
8	30	26	12	57	30					0	2
9	26	13	14	26	60					0	59,0
10	77	30	77	9	14					0	20,2
11	97	11	1	1	2	97	0			0	
12	96	12	1	1	0	0	96			0	
13	70	49	19	11	70	0	0			0	3
14	70	57	70	4	23	1	2			0	83,4
15	84	30	8	84	5	1	1			1	31,8
16	23	18	31	23	34	12				0	
17	78	32	3	4	15	78				0	
18	54	53	54	23	0	22				0	4
19	67	33	22	67	5	6				0	57,2
20	64	34	22	4	64	10				0	34,0
21	11	24									
22	60	25									
23	69	43									5
24	85	25									53,6
25	43	42									31,8
	62,76	30,56	average								
	22,88	12,90	st.dev.								

53,08	36,96	62,76	30,56	68,80	40,16	62,40	37,56
21,17	10,90	22,88	12,90	17,39	21,19	25,68	13,62

Figure 88 – Résultats à interpréter (Ibid.)

4.1.5 Exemple d'application

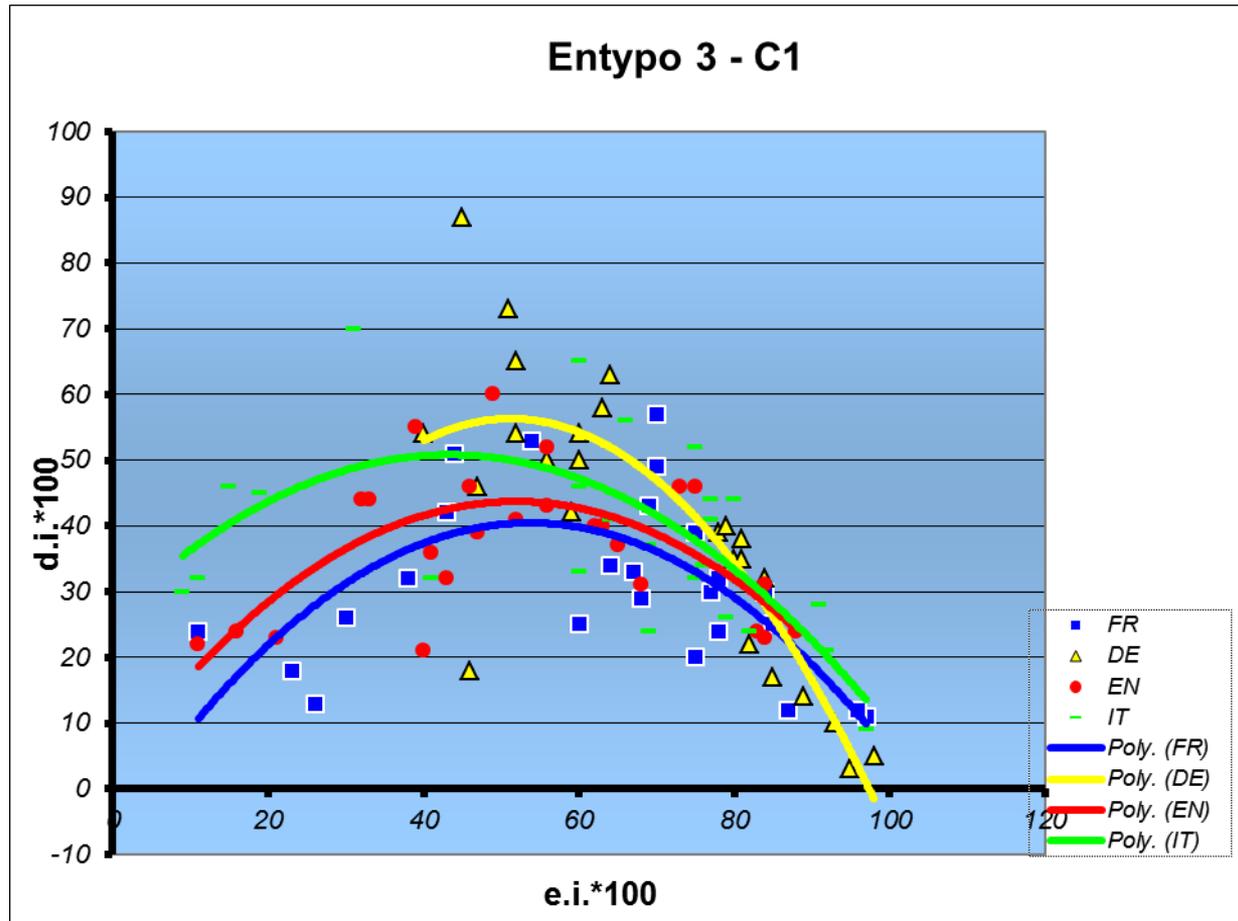


Figure 89 – Indices de discrimination (Ibid.)

4.1.5 Exemple d'application

CONSTATATION

L'indice de discrimination moyen (30,56) est inférieur à celui des mêmes épreuves dans les autres langues (moyenne des indices de discrimination des autres épreuves : 38,22)

HYPOTHÈSE

La présence, parmi les réponses proposées, de l'option « On ne peut pas le savoir » peut perturber les examinés.



4.1.5 Exemple d'application

DISCUSSION

Pour que cette hypothèse soit validée, il doit exister

- une corrélation entre les mauvais indices de discrimination et la présence de l'option « On ne peut pas le savoir » dans le QCM, d'une part,
- une absence de corrélation entre les bons indices de discrimination et la présence de l'option « On ne peut pas le savoir » dans le QCM, d'autre part.

Ce n'est pas le cas : dans les deux seules activités au sein desquelles l'option « On ne peut pas le savoir » peut être proposée, l'indice moyen de discrimination est

- de 28,5 pour les questions comportant l'option « On ne peut pas le savoir » dans les réponses, et
- de 28,7 pour les questions ne comportant pas cette option.



4.1.5 Exemple d'application

Un écart de 0,2 entre les indices moyens de discrimination n'est pas significatif.

De toute façon, lorsqu'on travaille sur des collections d'items aussi réduites (25 items dont 5 ROC), aucune conclusion ne devrait être directement tirée de l'absence ou de la présence de quelque corrélation



4.1.5 Exemple d'application

INTERPRÉTATION

L'indice de discrimination particulièrement bas ($<.20$) de cinq items (7, 9, 11, 12, 16) a fait baisser la qualité de l'épreuve. Il faut tout d'abord bien distinguer les items en fonction de leur indice de facilité :

Les items 11 et 12, par exemple, étaient exagérément faciles (>80) : respectivement 97 et 96% de réussite ! C'est cette facilité qui a induit la réduction de l'indice de discrimination : ces items étaient tellement faciles à résoudre que la plupart des examinés, dont les résultats à l'ensemble de l'épreuve étaient moins bons, y sont arrivés aussi. L'explication de cette facilité réside dans le fait que les réponses étaient quasi « dictées » par le responsable du CNC dans la séquence audio.



4.1.5 Exemple d'application

La direction générale supervise plusieurs directions. L'une des plus importantes est, d'abord, la direction du cinéma qui contrôle le service chargé de la diffusion des films en salles. C'est le service de la distribution. Vient ensuite celle qui a sous ses ordres le service des affaires juridiques auquel est rattaché le service des professions, service qui délivre, entre autres, les cartes d'identité professionnelles aux techniciens de la production cinématographique.



4.1.5 Exemple d'application

L'item 7 étant, lui aussi très facile (87% de réussite), on peut attribuer la faiblesse de son indice de discrimination à cette grande facilité. Ce n'est en tout cas probablement pas la présence de l'option « On ne peut pas le savoir » qui a pu faire baisser le degré de discrimination

- puisque l'indice de facilité est particulièrement élevé et
- que la proportion des examinés qui ont choisi l'option « On ne peut pas le savoir » est particulièrement faible (8%).

Les indicateurs permettant de répondre étaient d'ailleurs particulièrement évidents :

Mais je n'ai jamais regretté. Et toi ?

Non, au contraire ça m'a ouvert des portes !



4.1.5 Exemple d'application

L'item 16 semble par contre difficile. La distribution des réponses étant large et harmonieuse, on peut penser que beaucoup d'examinés ont répondu au hasard.

La séquence ne semblant pas particulièrement difficile à comprendre, on peut imaginer que – cette activité étant absolument nouvelle – ce premier item ait surpris les candidats. Les quatre items suivants leur ont d'ailleurs paru beaucoup plus faciles et sont devenus particulièrement discriminants.

Ce dernier constat devrait nous encourager à maintenir cette activité qui, si on met de côté l'item 16 et l'effet de surprise qu'il a probablement créé, semble constituer l'activité la plus discriminante de l'épreuve (38).



4.1.5 Exemple d'application

Enfin, l'item 9 n'a, quant à lui, visiblement pas « fonctionné » correctement. Il semble que la séquence

Non. Cette fois-ci, on reste en contact. À l'époque du collège Saint Paul, ni toi ni moi n'étions fans des emails, maintenant la communication est plus facile !

ne montrait pas de façon suffisamment explicite que les deux interlocuteurs se sont connus au collège.

Ou bien des examinés, pensant que les interlocuteurs ont très bien pu se connaître avant le collège, ont préféré la réponse « On ne peut pas le savoir ».

Il est vrai que si cette dernière option n'avait pas été proposée, les examinés auraient répondu plus facilement, contraints par exemple de choisir entre trois lieux. Mais il est vrai aussi que les indicateurs permettant de répondre étaient particulièrement implicites.



4.1.5 Exemple d'application

CONCLUSIONS

Au terme de cette brève analyse, il n'est **pas** apparu que l'option « On ne peut pas le savoir » doive être systématiquement proscrite de l'épreuve 3 du niveau C1.

Il est par contre apparu que les items 7, 11 et 12 étaient exagérément faciles et que l'item 9 était mal construit.

La surprise que peut provoquer chez l'examiné l'apparition d'un nouveau type d'activité explique probablement les valeurs extrêmes des indices de facilité et de discrimination de l'item 16.

Le type de l'activité 4 permet une bonne discrimination et peut être conservé.

Cette proscription serait d'ailleurs en contradiction avec la perspective adoptée de l'usage de la langue, mais ceci est un autre chapitre.



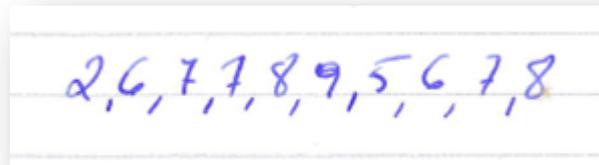
4.2 Analyse de la distribution des notes

- 4.2.1 Cas d'un contrôle en cours d'année
- 4.2.2 Cas d'un examen
- 4.2.3 Interprétation des biais



4.2.1 Cas d'un contrôle en cours d'année

Les notes de 10 élèves à un même contrôle sont les suivantes :



2, 6, 7, 7, 8, 9, 5, 6, 7, 8

Figure 90 – Scores de 10 élèves à un même contrôle



4.2.1 Cas d'un contrôle en cours d'année

Le **total** des notes obtenues s'élève à 65.

La **moyenne** des notes de la classe est égale à $65 / 10 = 6,5$

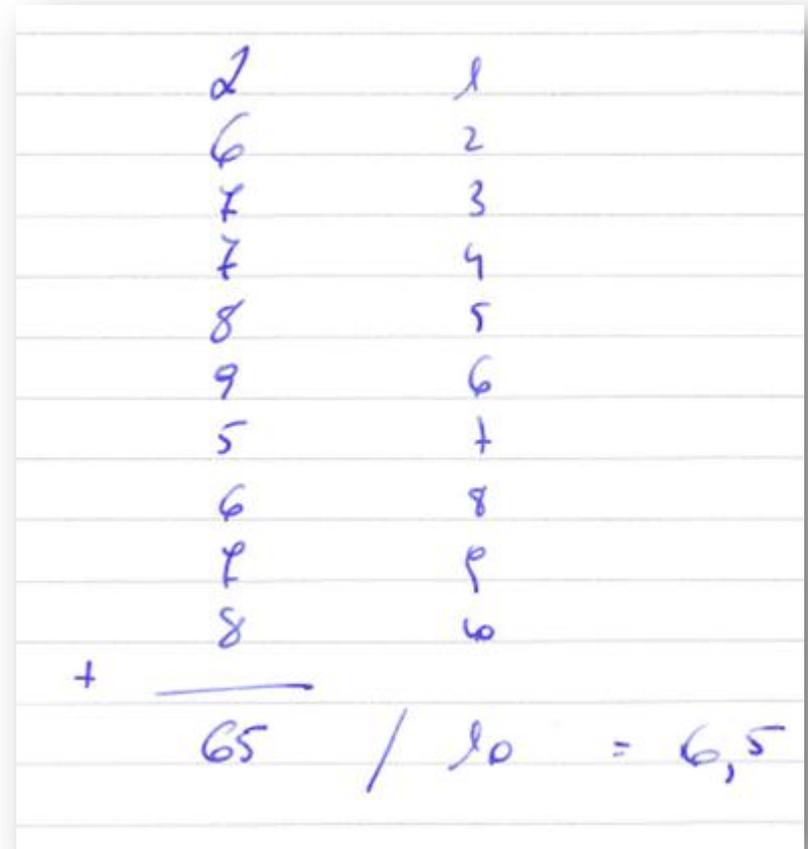


Figure 91 – Calcul des notes et de la moyenne



4.2.1 Cas d'un contrôle en cours d'année

Le calcul de cette moyenne peut apporter des informations précieuses. Ici, par exemple, la faiblesse de cette moyenne pourrait laisser penser :

- que la classe est encore loin d'avoir atteint l'objectif fixé (conclusion liée à l'organisation d'une évaluation critériée)
et/ou
- que le contrôle était vraiment trop difficile (conclusion liée à l'organisation d'une évaluation normative)

Ces informations restent des **hypothèses**.

Pour confirmer chacune d'entre elles, on peut toutefois dresser un petit tableau.



4.2.1 Cas d'un contrôle en cours d'année

Distribution des résultats

En **abscisse** (horizontalement), nous portons les notes obtenues à ce test.
En **ordonnée** (verticalement), nous indiquons le nombre d'élèves qui ont obtenu chacune de ces notes.

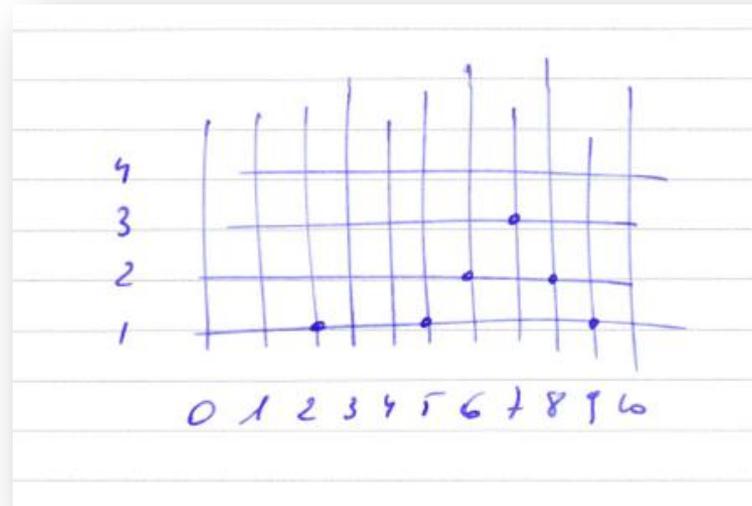


Figure 92 – Distribution des notes



4.2.1 Cas d'un contrôle en cours d'année

Ce tableau nous donne des informations supplémentaires.

Par exemple :

- Un élève se détache du lot. Peut-être n'avait-il pas étudié ?

ou encore

- Si on ne tient pas compte du résultat anormalement bas de l'élève qui a obtenu 2/10, la moyenne des résultats de la classe est finalement : $63 / 9 = 7$.



4.2.1 Cas d'un contrôle en cours d'année



Figure 93 – Distribution des notes

4.2.1 Cas d'un contrôle en cours d'année

Le contrôle n'était en tout cas ni trop facile, ni trop difficile puisque les notes se répartissent symétriquement autour de la moyenne.

C'est donc plutôt la classe qui n'a pas (encore) atteint l'objectif fixé.

Ici, par exemple, la faiblesse de cette moyenne pourrait laisser penser :

- que la classe est encore loin d'avoir atteint l'objectif fixé
et/ou
- que le contrôle était vraiment trop difficile.



4.2.1 Cas d'un contrôle en cours d'année

Si nous pratiquons une évaluation normative : nous avons maintenant l'assurance d'avoir organisé un contrôle qui correspondait au « niveau » de la classe.

Si nous pratiquons une évaluation critériée : nous savons maintenant que l'objectif qui en constitue le critère n'est atteint qu'à 70%.

Nous tirerions d'autres conclusions si notre petit tableau avait ressemblé à ceci :



4.2.1 Cas d'un contrôle en cours d'année

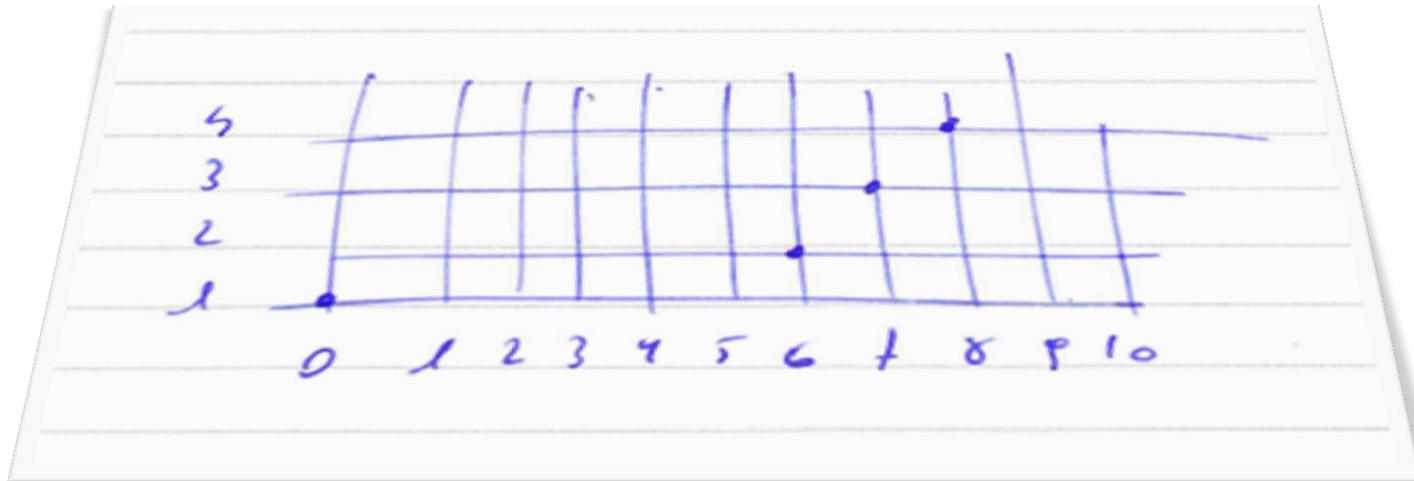


Figure 94 – Autre distribution des notes



4.2.1 Cas d'un contrôle en cours d'année

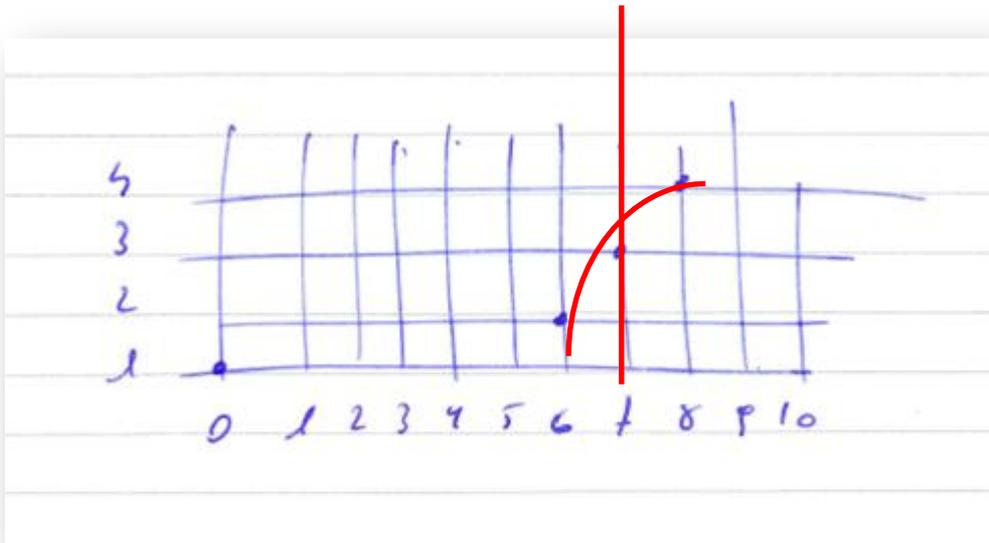


Figure 95 – Distribution anormale des notes

Ici, même si nous ignorons le zéro obtenu par un élève qui avait, par exemple, triché pendant le contrôle, les notes ne sont pas réparties d'une façon symétrique autour de la valeur centrale (7) : beaucoup de 8, peu de 6.

Cela indique que le contrôle était trop facile pour la classe.



4.2.1 Cas d'un contrôle en cours d'année

Il reste à savoir pourquoi ce contrôle s'est révélé trop facile :

- Certaines questions étaient-elle un peu stupides ?

En tout cas, pas toutes, parce que sinon on aurait eu une moyenne plus élevée, on aurait eu des 9 et des 10.

- Les élèves viennent-ils de faire d'énormes progrès ?

Mais il en reste encore à faire, parce que sinon on aurait eu une moyenne plus élevée, on aurait eu des 9 et des 10.



4.2.2 Cas d'un examen

Prenons un second exemple. Celui d'anciens examens du PALSO : 1654 candidats ont participé à un examen du niveau A2.

Voici une représentation graphique de la distribution de leurs notes totales.

Le tableau se lit comme les précédents. Il indique, par exemple que 250 candidats ont obtenu une note située entre 75 et 80.



4.2.2 Cas d'un examen

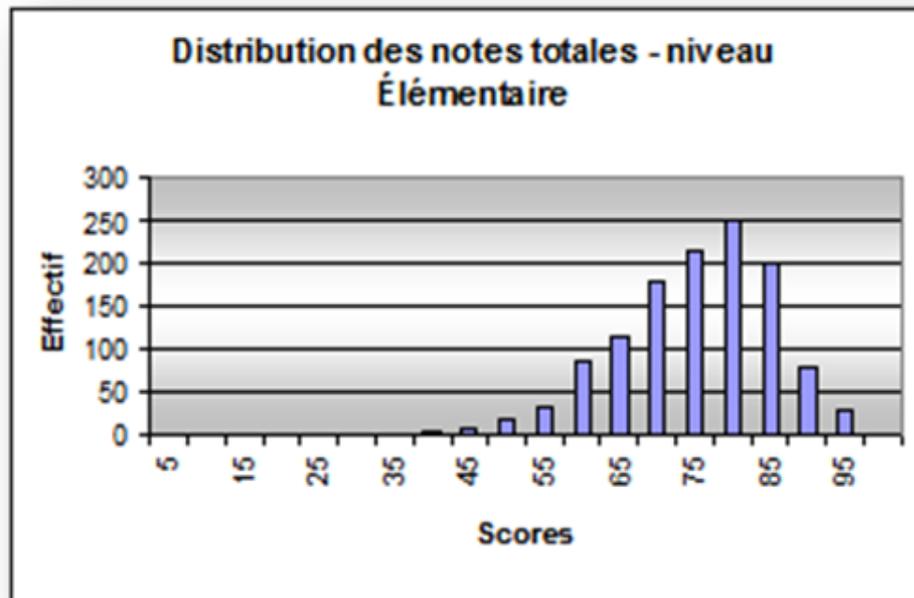


Figure 96 – Distribution des notes totales (Delhaye, 2001)



4.2.2 Cas d'un examen

La note moyenne qu'ils ont obtenue est de 73/100. À priori, l'examen pourrait donc sembler assez difficile. Pourtant ce n'est pas le cas :

- La courbe des résultats est légèrement biaisée vers la droite. Donc, les questions sont un peu trop faciles.
- Le seuil de réussite est de 60% : seulement 11,8 % des candidats ne recevront pas de certificat. La signification de la certification octroyée est donc très faible.



4.2.2 Cas d'un examen

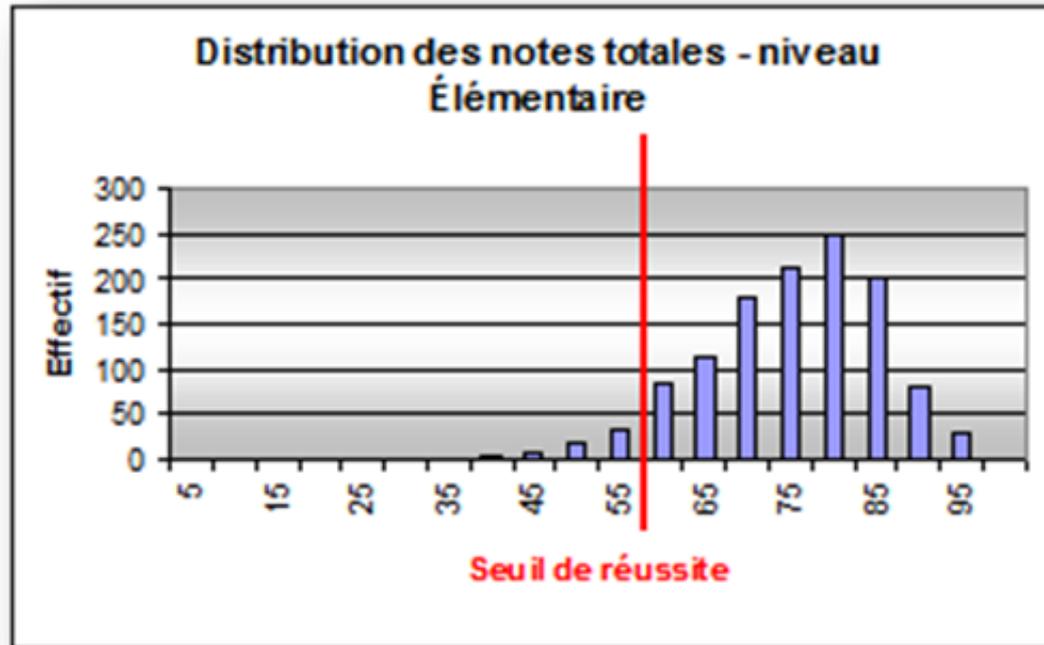


Figure 97 – Seuil de réussite (Delhay, 2001)



4.2.2 Cas d'un examen

Une autre explication à ce haut degré de facilité de réussite peut être que pour des raisons politiques et/ou commerciales, le seuil de réussite de l'épreuve a été placé bas.

En effet, au plan politique, les commanditaires peuvent considérer l'obtention du diplôme élémentaire comme un encouragement pour l'entreprise d'études visant l'obtention du certificat. Au plan commercial, il est certain que la population des élèves se dirigera plus volontiers vers la préparation d'un diplôme facile à obtenir.

Une mention (très bien, bien, assez bien, passable) figure certes sur les certifications octroyées, mais rien ne permet de les interpréter.



4.2.2 Cas d'un examen

Un employeur pourrait donc être trompé par la présentation d'une certification niveau élémentaire, mention passable, qui, en définitive, signifie pauvrement que son détenteur ne fait pas partie des 11% d'examinés qui ne sont pas arrivés à l'obtenir.

Tout ceci pose le problème de la **sémiologie** du score dont la définition doit être soigneusement établie par les commanditaires de l'examen.



4.2.3 Interprétation des biais

La représentation graphique de la **distribution des notes** obtenues à un test ressemble le plus souvent une courbe en forme de cloche. Elle est dite « normale ».

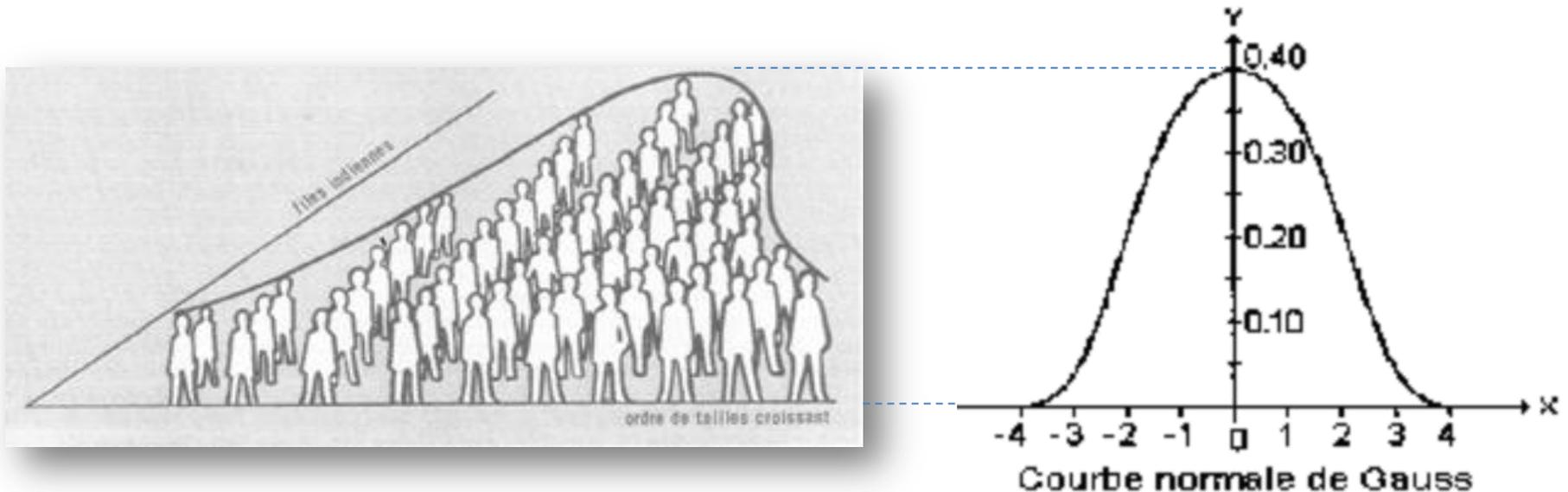


Figure 98 – Loi normale et courbe de Gauss
(SÉRÉBRIAKOFF V. & LANGE St., 1979)

4.2.3 Interprétation des biais

Si la cloche penche vers la droite, le test comporte des items trop faciles.

Si la cloche penche vers la gauche, le test comporte des items trop difficiles.

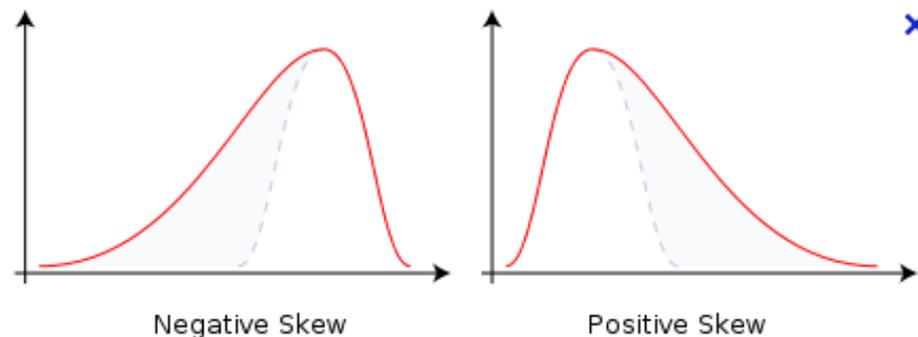


Figure 99 – Biais dans la distribution des notes (Ibid.)

4.3 Établissement de corrélations

4.3.1 Accord interjuge

4.3.2 Exemple d'application



4.3 Établissement de corrélations

Il est possible de tirer des renseignements utiles de la mise en corrélation de certaines notes. C'est le cas de l'estimation de l'accord interjuge.

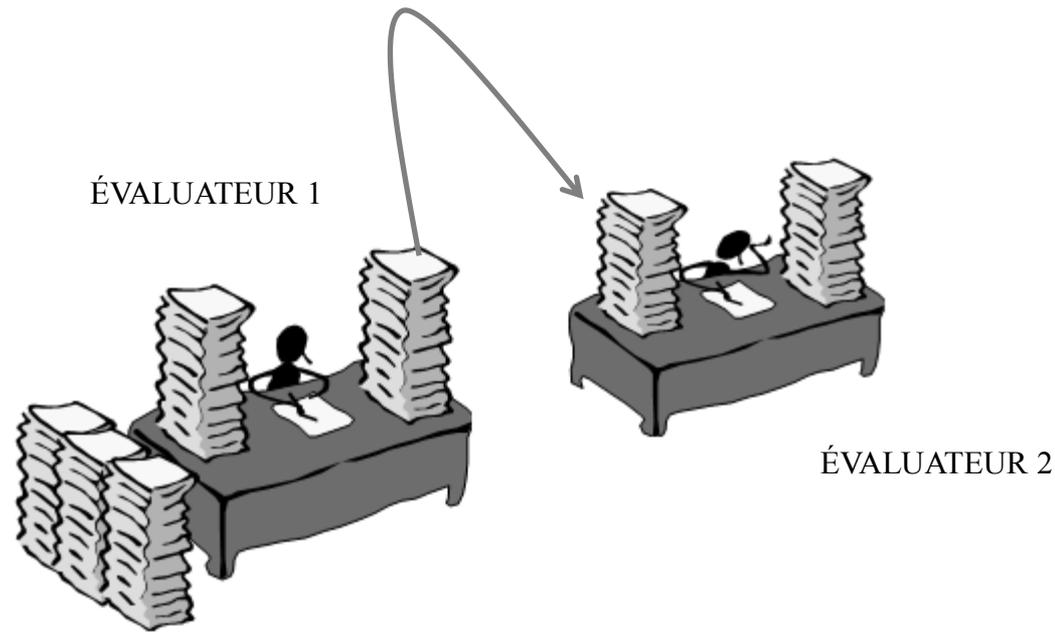


Figure 100 – Nécessité d'un accord interjuge



4.3.1 Accord interjuge

Quand on n'évalue pas des réponses fermées (QCM), on doit organiser des doubles corrections.

On peut comparer les notes obtenues en traçant des droites de régression.

Voici les notes attribuées à une même production écrite par deux examinateurs.



4.3.1 Accord interjuge

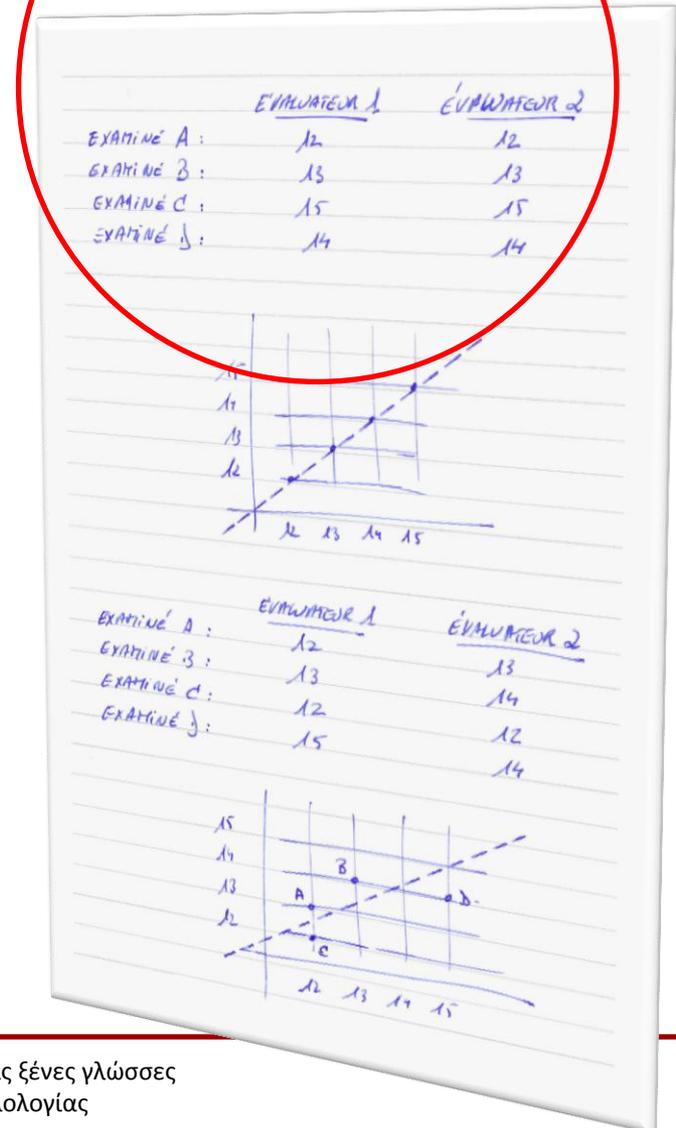


Figure 101 – Notes attribuées à une même production écrite par deux examinateurs



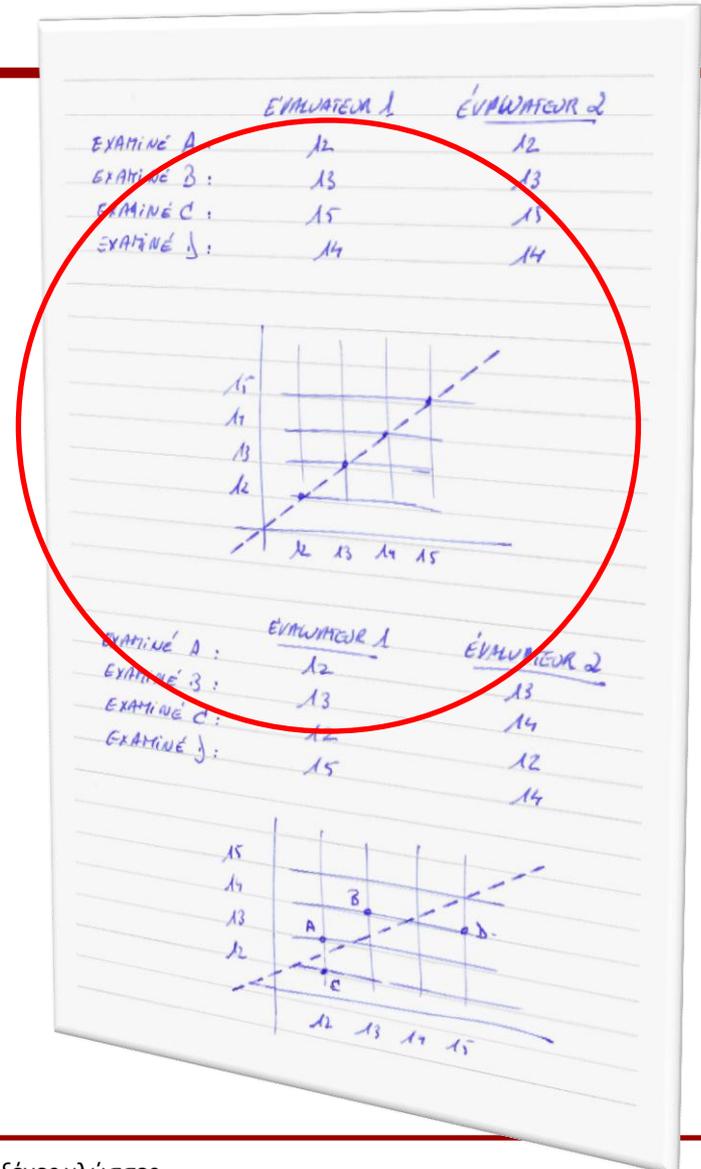
4.3.1 Accord interjuge

Plaçons ces notes sur le graphe :

- en abscisse, les notes de l'évaluateur 1 ;
- en ordonnée, les notes de l'évaluateur 2.

Traçons une droite qui réunit tous ces points

Figure 102 – Figuration des notes



4.3.1 Accord interjuge

Cette droite passe par les points de coordonnées (12,12) (13,13) (14,14) etc.

Ce cas de figure est idéal. En réalité, les examinateurs ne donne pas toujours la même note à un même candidat.

Considérons cette nouvelle série de notes, plus réalistes.

Plaçons les points (12,13) (13,14) (12,12) etc. sur le graphe.

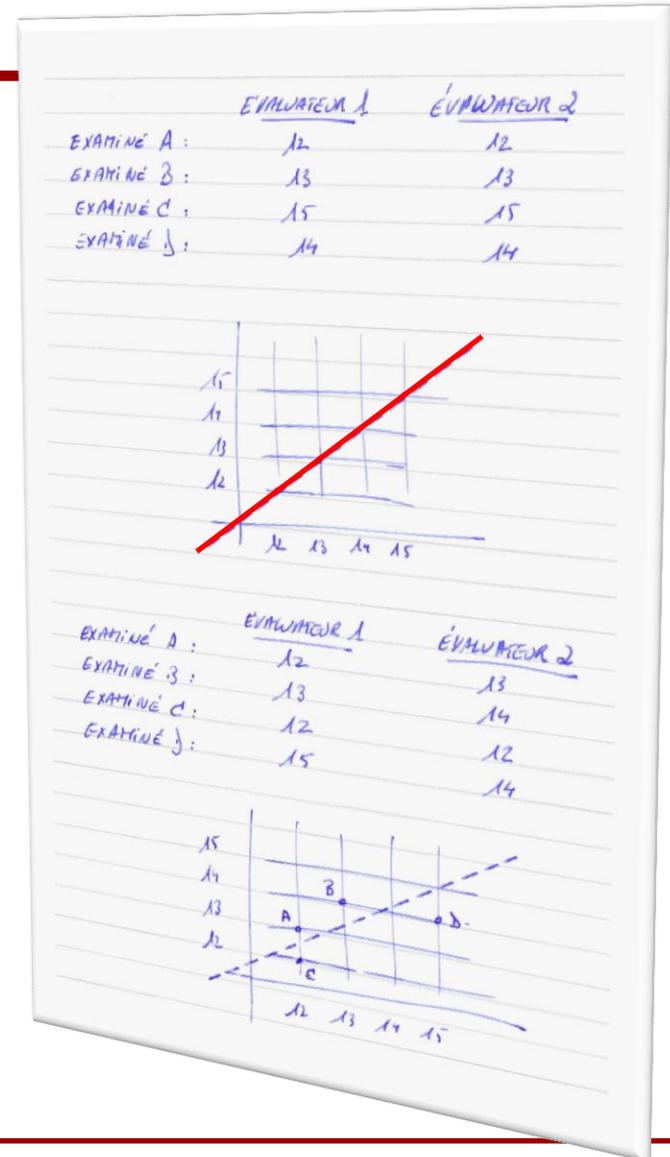


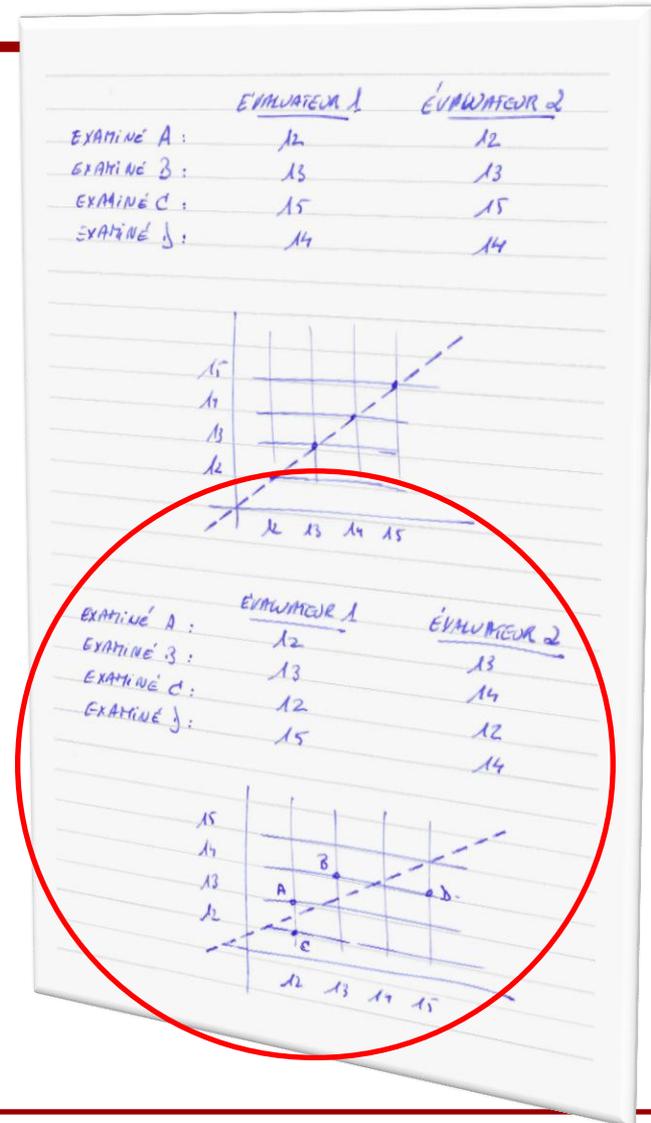
Figure 103 – Droite de régression



4.3.1 Accord interjuge

L'inclinaison de la droite de régression (tendance) a changé et les points ne se trouvent plus exactement dessus.

Figure 104 – Autres notes



4.3.1 Accord interjuge

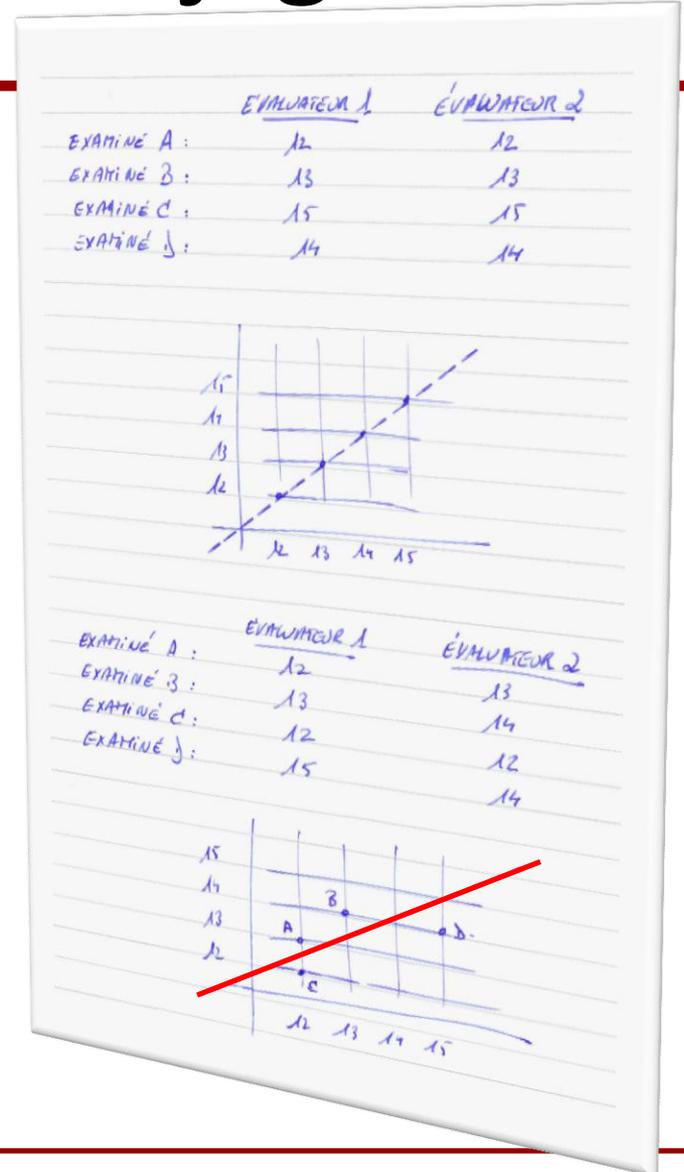


Figure 105 – Droite de régression



4.3.2 Exemple d'application

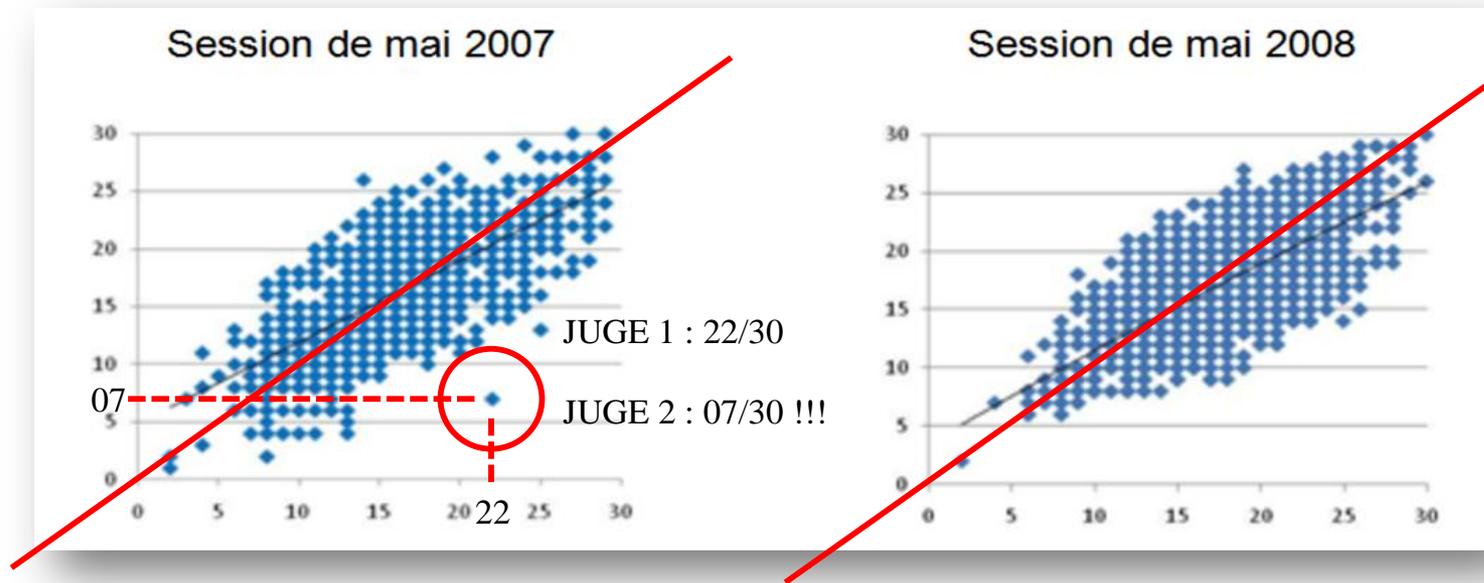


Figure 106 – Droites de régression et nuages (Delhayé 2009)



4.3.2 Exemple d'application

Voici par exemple la corrélation entre les notes interjuge à deux sessions des examens du ΚΠγ. Il est entendu que chaque point peut cacher un empilement de points.

On remarque :

- que les notes ne se trouvent pas toutes sur la droite attendue. Il y a donc des écarts entre les évaluations des deux juges (Figure 106)
- que la droite de régression ne passe pas par (0,0) et par (30,30). Un juge est plus sévère avec les examinés faibles et plus généreux avec les forts (Figure 107).
- que les points sont plus épars en 2007. La grille d'évaluation était peut-être mieux faite en 2008 (Figure 108).
- que les points forment un nuage en amande en 2008 (Figure 109). Les juges sont plus d'accord qu'avant sur les notes à donner, surtout quand elles sont extrêmes.



4.3.2 Exemple d'application

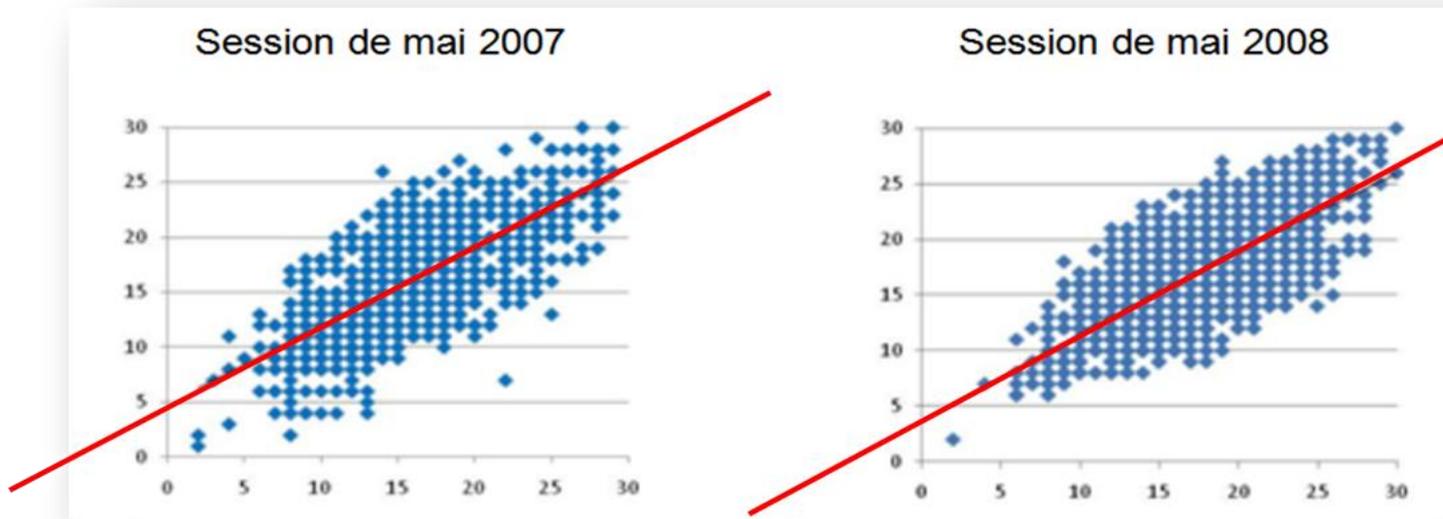


Figure 107 – Droites de régression (Ibid.)



4.3.2 Exemple d'application

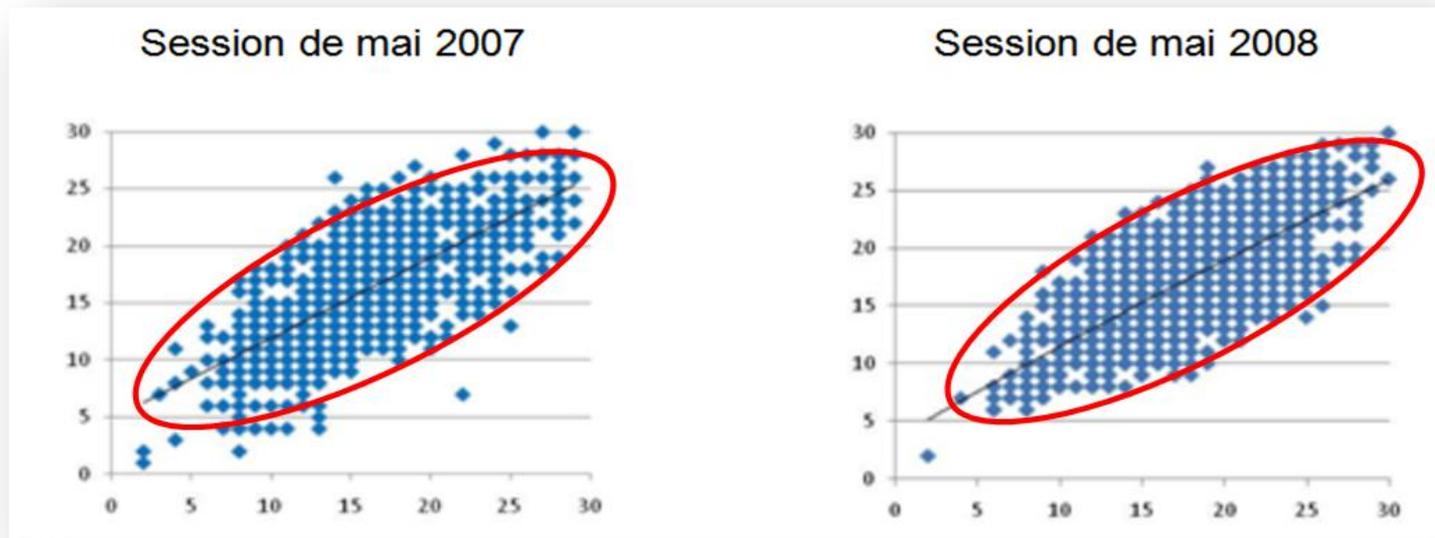


Figure 108 – Nuages (Ibid.)



4.3.2 Exemple d'application

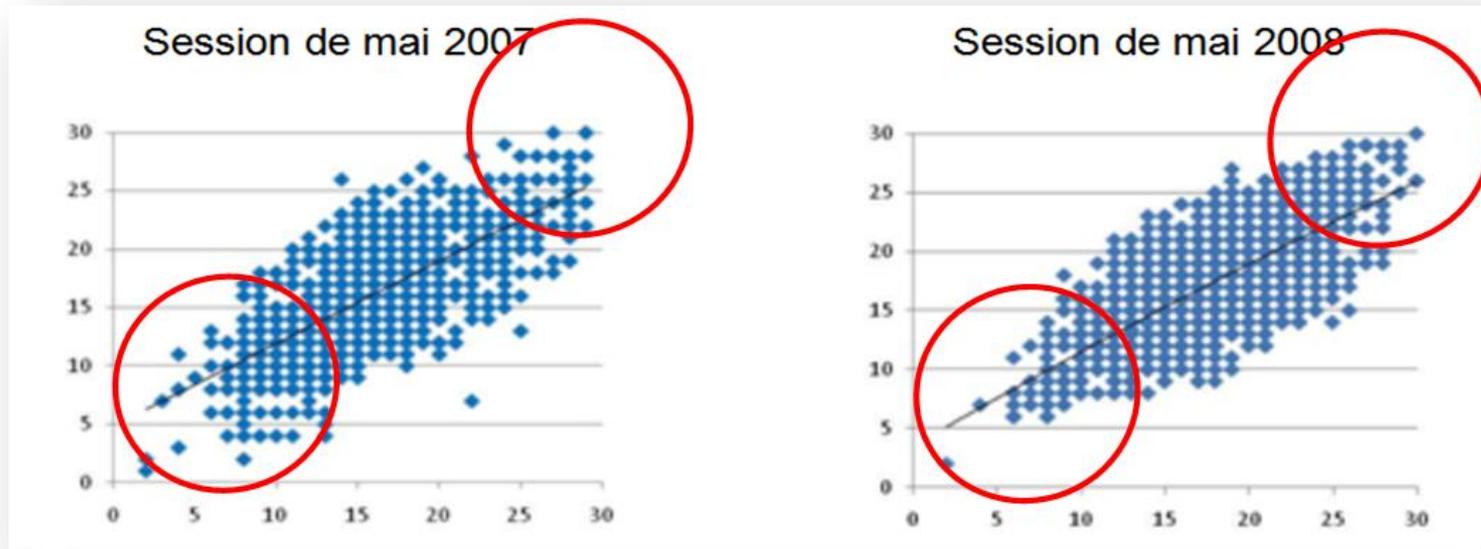


Figure 109 – Aplatissement d'un nuage (Ibid.)



4.3.2 Exemple d'application

Nous n' avons étudié que l'accord interjuge.

D'autres corrélations peuvent être calculées et représentées de cette même manière :

- constance dans le temps d'un même examen,
- équivalence entre deux examens appartenant à des dispositifs différents,
- etc.



Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Olivier Delhaye.
«Évaluation des compétences de communication
en langue étrangère». Έκδοση: 1.0. Θεσσαλονίκη 2014. Διαθέσιμο από τη
δικτυακή διεύθυνση: <http://eclass.auth.gr/courses/OCRS422/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Παρόμοια Διανομή [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

[1] <http://creativecommons.org/licenses/by-sa/4.0/>





Τέλος Ενότητας

Επεξεργασία: Ανδρομάχη Μουρτζούχου
Θεσσαλονίκη, 30 Ιουνίου 2015



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

