

LIBIRT / EIRT: programme «open-source» pour la théorie de réponses aux items

Abdous, B.^{1,2}, Valois, P.³, & Germain, S.³

¹Haut Commissariat au Plan, Rabat, Morocco

²Université Mohamed VI Polytechnique, Benguerir, Morocco

³Université Laval, Québec, Canada

**L'évaluation en éducation et en formation:
approches,
enjeux et défis**

Rabat, 22-23 octobre 2015



PLAN

- [Psychométrie](#)
- [Théorie de réponses aux items](#)
- [Programme Libirt](#)

INTRODUCTION : PSYCHOMÉTRIE

- Domaine d'étude qui s'intéresse aux théories de la mesure en éducation et en psychologie
- Mesure des connaissances, des habiletés, des attitudes, des traits de la personnalité, etc.

INTRODUCTION : PSYCHOMÉTRIE

- Construction et validation d'outils d'évaluation tels que :
 - Questionnaires
 - tests
 - Jugements d'experts
 - Tests de personnalité
 - etc.

INTRODUCTION : HISTOIRE DE LA PSYCHOMÉTRIE¹

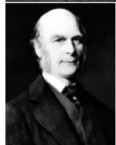
Période victorienne : Darwin, Galton, Cattell, Pearson, ...



Ch. Darwin
(1859)
*The Origin
of Species*



J. Mc.
Cattell
(1860-1944)
Mental tests



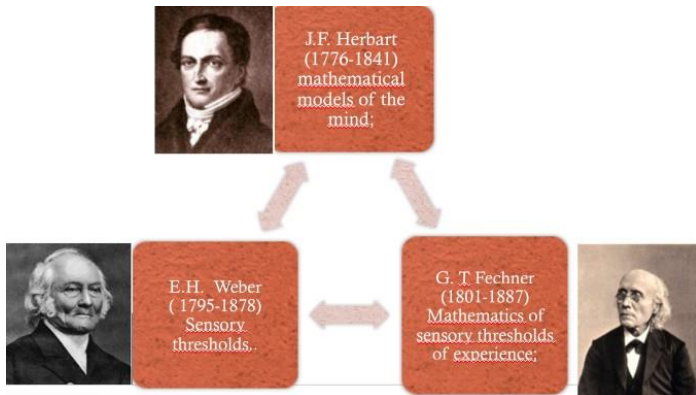
F. Galton
(1869)
*Hereditary
Genius*



K. Pearson
(1857-1936)
*Statistical
regression*

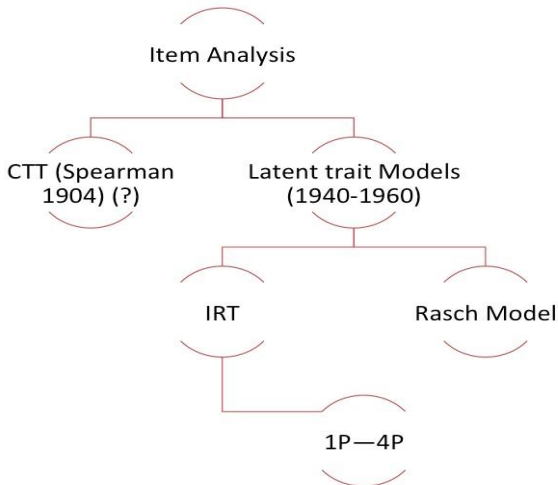
INTRODUCTION : HISTOIRE DE LA PSYCHOMÉTRIE²

Période allemande : Herbart, Weber, Fechner, ...



INTRODUCTION : THÉORIES DE LA MESURE

- Les psychométriciens ont développé plusieurs théories de la mesure



THÉORIE CLASSIQUE DES TESTS (CTT)

La CTT assume que le score brut (X) obtenu par une personne est formé de deux composantes

$$X = V + \varepsilon$$

où



V = score vrai de la personne (*considéré comme le score moyen que la personne obtiendrait suite à un nombre infini de passations*)



ε = erreur aléatoire (non observée).

Postulats : $\varepsilon \sim N(0, \sigma^2)$ and $\rho(V, \varepsilon) = 0$.

THÉORIE DES RÉPONSES AUX ITEMS (TRI)

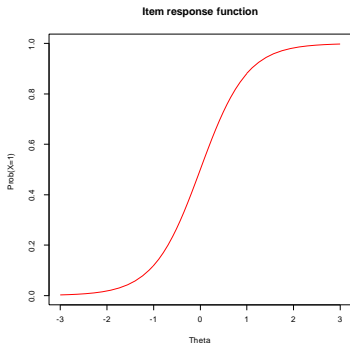
La TRI constitue un cadre pour évaluer la qualité des items et des tests.

Elle est basée sur trois postulats principaux

- Un concept unidimensionnel nommé θ ;
- L'indépendance local des items ;
- La réponse d'une personne à un items peut être modélée par une fonction mathématique de réponse aux items

LA FONCTION DE RÉPONSES AUX ITEMS

TRI : un paradigme pour le développement, l'analyse et la mesure de concepts réalisée avec des tests, des questionnaires, des entretiens, etc.



- θ : Habileté de l'étudiant/personne's sur l'axe x
- Probabilité d'une bonne réponse à **un** item d'un test sur l'axe y

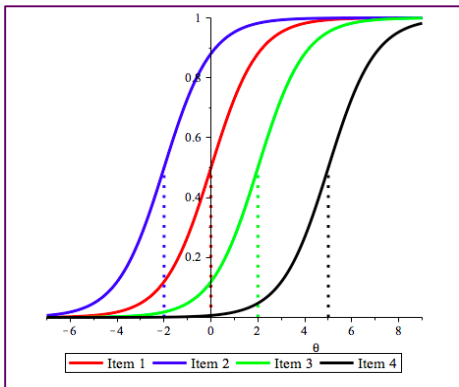
THÉORIE DE RÉPONSE AUX ITEMS : 1PL

$$Pr(X_i = 1 | \theta) = 1 / (1 + e^{-(\theta - b_i)})$$

- b_i = Paramètre de difficulté
- b_i indique si l'item est facile ou difficile

THÉORIE DE RÉPONSE AUX ITEMS : 1PL

$$Pr(X_i = 1 | \theta) = 1 / (1 + e^{-(\theta - b_i)})$$



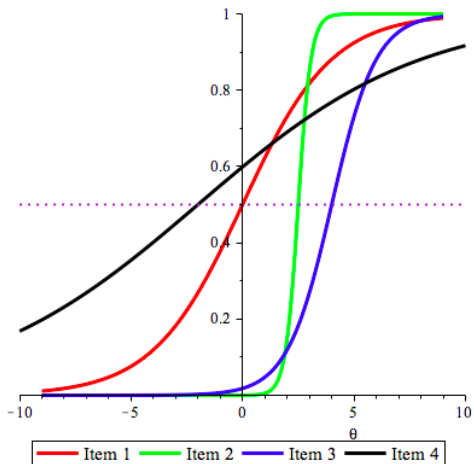
THÉORIE DE RÉPONSE AUX ITEMS : 2PL

$$Pr(X_i = 1 | \theta) = 1 / (1 + e^{-a_i(\theta - b_i)})$$

- a_i = Paramètre de discrimination
- a_i indique la puissance d'un item pour discriminer entre les élèves qui ont fait beaucoup d'apprentissage et ceux qui en ont fait moins
- b_i = Paramètre de difficulté

THÉORIE DE RÉPONSE AUX ITEMS : 2PL

$$Pr(X_i = 1 | \theta) = 1 / (1 + e^{-a_i(\theta - b_i)})$$



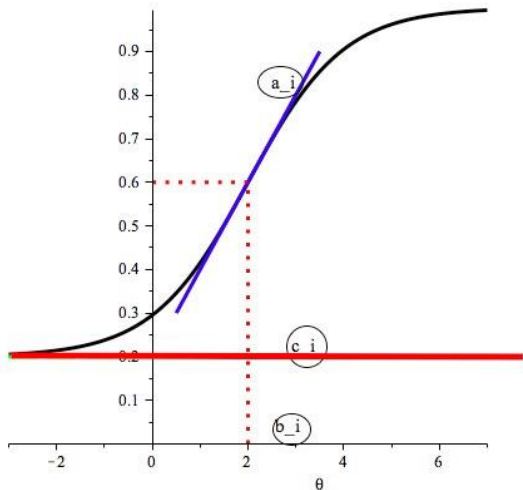
THÉORIE DE RÉPONSE AUX ITEMS : 3PL

$$Pr(X_i = 1 | \theta) = c_i + (1 - c_i) / (1 + e^{-a_i(\theta - b_i)})$$

- a_i = Paramètre de discrimination
- b_i = Paramètre de difficulté
- c_i = Paramètre de pseudo-chance

THÉORIE DE RÉPONSE AUX ITEMS : 3PL

$$Pr(X_i = 1 | \theta) = c_i + (1 - c_i) / (1 + e^{-a_i(\theta - b_i)})$$



EIRT : UN « ADD-IN » D'EXCEL

- EIRT : est un « add-in » *Excel* pour la TRI
 - Réponses binaires
 - Réponses graduées
 - Réponses nominales
- **Estimateurs des paramètres** : maximum de vraisemblance marginale
- **Estimateurs non paramétriques**: estimateurs à noyau et par maximum de vraisemblance marginale pénalisée
- **Estimation de traits latents**: estimateur de moyenne a posteriori

EIRT : EXCEL ADD-IN

- Le programme *eirt* peut être téléchargé à l'adresse suivante :
<http://sourceforge.net/projects/libirt/files/eirt>
- Le nom du programme d'installation prend la forme **eirt-x.x.x.exe**, où **x.x.x** désigne le numéro de version.

ÉTAPE 1 : SÉLECTION DES DONNÉES

The screenshot displays the Microsoft Excel interface with a dialog box titled "eirt" open. The dialog box is titled "Step 1 of 4" and contains the following text: "Select the range containing the data to analyse." Below this text is a text box containing the selection range "\$A\$1:\$K\$1001". There are three checkboxes with labels: "The subjects' label are in the first column of the selection", "The items' label are in the first line of the selection", and "The correction key is in the first line of the selection". At the bottom of the dialog box are four buttons: "< Back", "Cancel", "Help", and "Next >".

	A	B	C	D	E	F	G	H	V8
1		V1	V2	V3	V4	V5	V6	V7	
2	1	1	0	1	0	0	0	0	0
3	2	1	0	1	1	1	1	1	0
4	3								
5	4								
6	5								
7	6								
8	7								
9	8								
10	9								
11	10								
12	11								
13	12								
14	13								
15	14								
16	15								
17	16								
18	17								
19	18								
20	19								
21	20								
22	21								
23	22								
24	23								
25	24								
26	25								
27	26								
28	27								
29	28								
30	29								
31	30								
32	31								
33	32								
34	33								

ÉTAPE 2 : SÉLECTION DU TYPE DE DONNÉES

Step 2 of 4

Select the type of data to analyse and specify the format.

Dichotomous

Multiple choice

Graded

Mixed

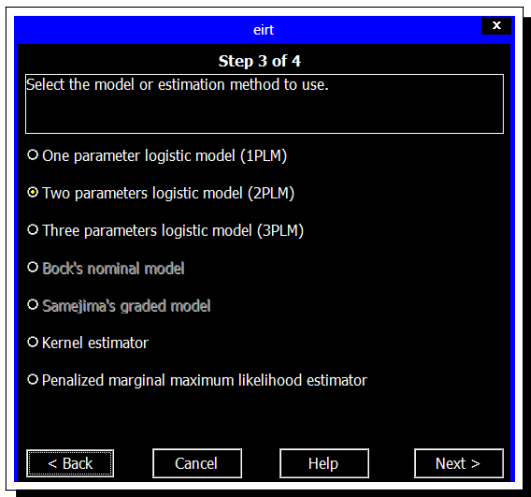
Success

Missing value

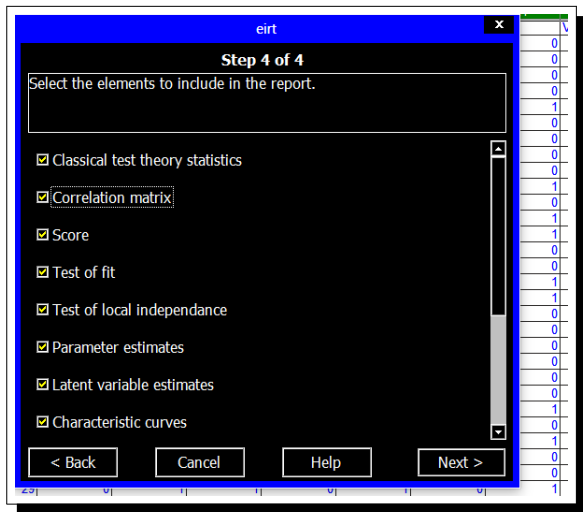
Answer key 0 Items

< Back Cancel Help Next >

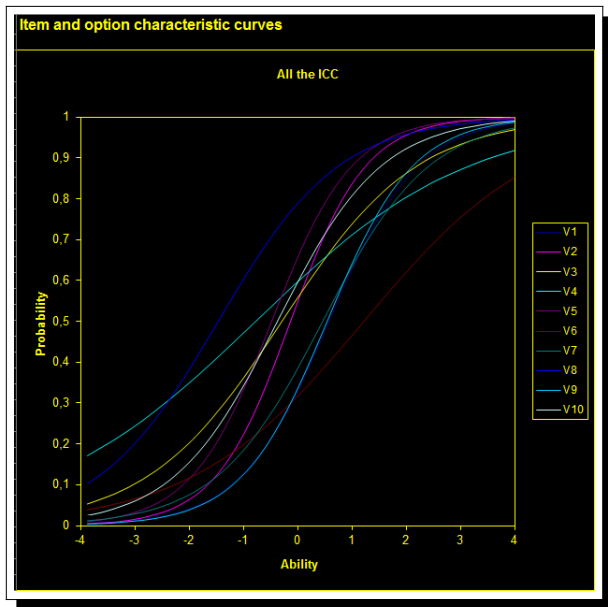
ÉTAPE 3 : CHOIX DU MODÈLE



ÉTAPE 4 : OPTIONS DU RAPPORT



RÉSULTATS: COURBES CARACTÉRISTIQUES D'ITEMS (ICC)



RÉSULTATS : TCC

Classical test theory statistics

<u>Number of subject</u>	<u>Number of item</u>	<u>Number of missing response</u>	<u>Score's mean</u>	<u>Score's standard deviation</u>	<u>Cronbach's Alpha</u>
1000	10	0	5,096	2,378	0,653

<u>Item</u>	<u>Mean (difficulty)</u>	<u>Standard deviation</u>	<u>Correlation (discrimination)</u>
V1	0,366	0,482	0,360
V2	0,533	0,499	0,407
V3	0,550	0,497	0,279
V4	0,591	0,492	0,188
V5	0,617	0,486	0,396
V6	0,330	0,470	0,221
V7	0,404	0,491	0,333
V8	0,756	0,429	0,260
V9	0,371	0,483	0,385
V10	0,578	0,494	0,338

Estimation summary

Data source Range '\$B\$2:\$K\$1001' in sheet 'data' of file 'Binary.xls'

Method Baye's modal estimator (BME)

Model Two parameters logistic model (2PLM)

No item were ignored by the estimation process.

All the items converged.

Test of fit

<u>Item</u>	<u>Chi-square</u>	<u>Degrees of freedom</u>	<u>P-Value</u>
V1	0,310	10	1,000

APPLICATIONS : POINT DE COUPURE

Fig. 1: point de coupure = 0

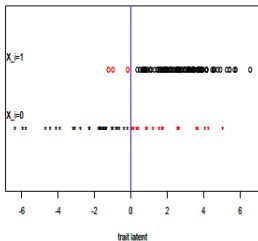


Fig. 2: point de coupure = 1.5

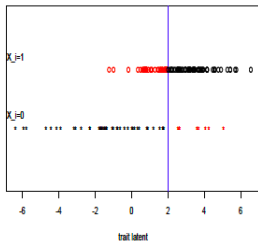


Fig. 3: point de coupure = -5

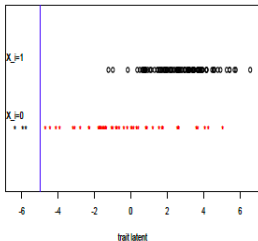
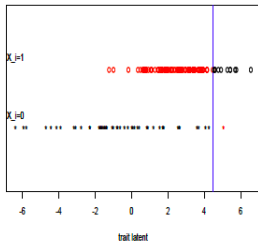


Fig. 4: point de coupure = 4.5



APPLICATIONS : POINT DE COUPURE

Fig. 1: point de coupure = 0

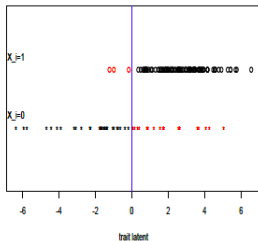


Fig. 2: point de coupure = 1.5

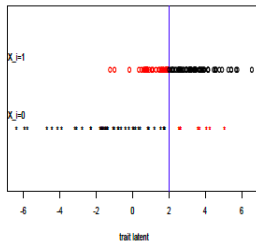


Fig. 3: point de coupure = -5

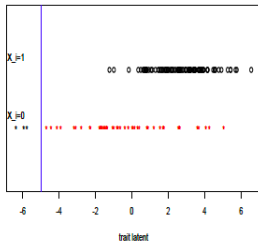
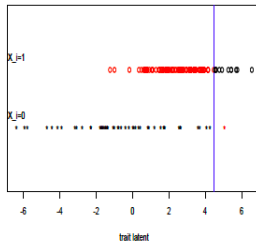


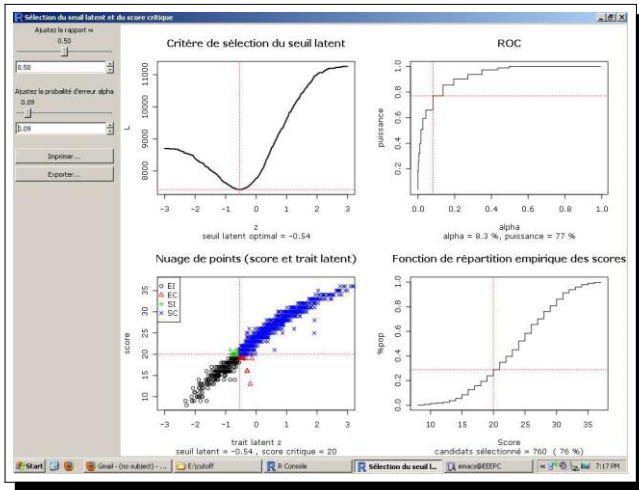
Fig. 4: point de coupure = 4.5



APPLICATIONS : POINT DE COUPURE

$$N \sum_{i=1}^n \left[w_1 \int_{-\infty}^z P_i(u) f(u) du + w_2 \int_z^{\infty} (1 - P_i(u)) f(u) du \right].$$

APPLICATIONS : POINT DE COUPURE



L'évaluation formative en français : analyse métrologique

Pierre Valois
Erick Falardeau
Frédéric Guay
Belkacem Abdous

COLLOQUE INTERNATIONAL
L'EVALUATION EN EDUCATION
ET EN FORMATION
APPROCHES, ENJEUX ET DÉFIS



PROGRAMME

RABAT, 22-23 OCTOBRE 2015

État de la situation

- L'évaluation de la lecture à l'école secondaire au Québec est de plus en plus préoccupante (Falardeau, Guay, & Valois, 2012)

État de la situation

- Au Québec, selon le programme de français au secondaire, les élèves doivent développer la compétence de « lire et apprécier des textes variés »
- L'élève est invité à « fonder une appréciation critique de textes littéraires et courants » et est évalué selon des critères (MEQ, 2003, p. 98)

État de la situation

- Or ces critères d'évaluation sont définis de façon trop globale pour permettre l'élaboration d'outils efficaces pouvant favoriser l'autorégulation des élèves en lecture

Critères d'évaluation de la compétence

- Compréhension des éléments significatifs d'un ou de plusieurs textes
- Justification pertinente des réactions à un ou plusieurs textes
- Recours à des stratégies de lecture appropriées

Constats scientifiques

- De nombreux chercheurs ont lié l'apprentissage de la lecture (Giasson, 1995) à une évaluation continue, fréquente et explicite
- L'évaluation devient alors partie prenante du processus d'apprentissage et requiert des outils clairs, qui accroissent le contrôle de l'élève sur ses tâches de lecture

Constats scientifiques

- Pour qu'un élève consolide sa compétence à lire et apprécier et maîtrise tous les processus qu'elle implique, il doit disposer, entre autres, d'outils d'évaluation formative qui proposent un découpage clair des différentes composantes de la compétence

État de la situation

- Au vue de tout ceci, comment donc assurer la qualité métrologique de l'évaluation formative en lecture au secondaire ?

Objectif

Développer un outil d'évaluation formative, en lecture au secondaire, et en tester la fidélité

Méthode

- 602 élèves qui proviennent du premier cycle du niveau secondaire
 - ✓ $n = 279$ pour le premier secondaire
 - ✓ $n = 323$ pour le deuxième secondaire
- Moyenne d'âge = 13,1 ans

Texte à lire

Gazaille, M.-P. (2002). Adolescents, publicité et surconsommation. Parents ados, 3(3), 5-6.

Omniprésente dans notre quotidien, la publicité informe, suggère et, de plus en plus, dicte ce qui est « in » ou « out ». Magazines, affiches, pauses publicitaires, elle s'affirme un peu plus chaque jour et nous suit où que l'on aille. Mais comment réagir lorsque ce sont nos ados qui y sont confrontés et qui doivent composer avec ces « dictats » et cette influence constante?

« Lire et apprécier des textes variés »



1. Je construis du sens
2. Je porte un jugement critique
3. Je réfléchis à ma pratique de lecteur
4. Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

Composante 1 : Je construis du sens

- Je planifie ma lecture
- Je comprends un texte
- J'interprète un texte
- Je réagis à un texte

Exemple d'items (réagir à un texte)

**Selon toi, les compagnies manipulent-elles à ce point les adolescents par la publicité ?
Justifie ta réponse en t'appuyant à la fois sur des éléments d'information du texte et sur ton expérience personnelle.**

Composante 1 : Je construis du sens (réagir à un texte)

5 POINTS

- Pour justifier son opinion, l'élève s'appuie clairement sur un élément du texte et sur un élément de son expérience personnelle, et le lien entre les deux est approprié.

Exemple : *Oui, je crois que les adolescents sont manipulés. Les publicitaires font exprès de rendre les nouveaux produits attrayants en faisant toutes sortes de promesse, comme les cellulaires. il est vrai qu'on s'en lasse rapidement. À mon école, tout le monde en a un parce que c'est la tendance. Pourtant, plusieurs élèves ne s'en servent même pas.*

Composante 1 : Je construis du sens (réagir à un texte)

4 POINTS

- L'élève justifie son opinion, mais un des deux éléments de réponse ou le lien tracé entre eux est plus ou moins pertinent.

Exemple : Oui, je crois que les adolescents sont manipulés par les compagnies, car elles leur font croire que leurs produits sont essentiels. Je connais quelqu'un qui est tombé dans le panneau. Il s'est acheté un iPod Touch

Composante 1 : Je construis du sens (réagir à un texte)

3 POINTS

- Pour justifier son opinion, l'élève s'appuie de façon plus ou moins pertinente sur un élément du texte, mais pas sur son expérience personnelle.

Exemple : Oui, je crois que les adolescents sont manipulés par les compagnies, car elles font de la publicité partout, dans les magazines, pendant les émissions de télévision, et même dans les écoles

Composante 1 : Je construis du sens (réagir à un texte)

2 POINTS

- L'élève ne s'appuie pas sur un élément du texte pour justifier son opinion, même si l'élément en lien avec son expérience personnelle est pertinent.

Exemple : *Non, je ne crois pas que nous sommes manipulés. Les jeunes de ma classe n'ont pas tous le dernier iPod.*

Composante 1 : Je construis du sens (réagir à un texte)

1 POINT

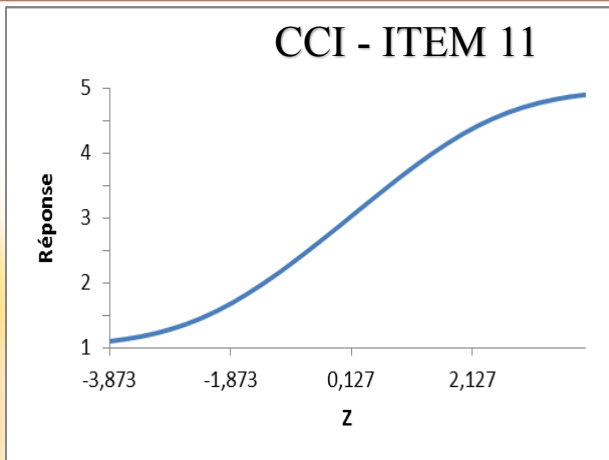
- L'élève justifie son opinion à l'aide d'éléments textuels ou personnels non pertinents.

Exemple : *Oui, je crois que les adolescents sont manipulés par les compagnies, car les vendeuses sont énervantes avec leur vente à pression.*

Je construis du sens (réagir à un texte)

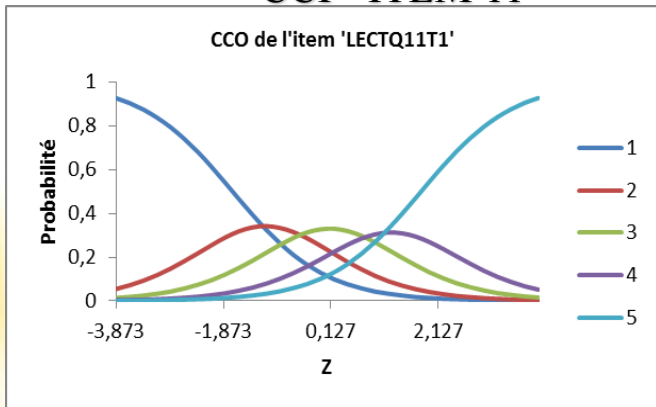
Score	Fréquence	%
1	102	16,9
2	132	21,9
3	159	26,4
4	117	19,4
5	92	15,4
Total	602	100,0

Je construis du sens (réagir à un texte)



Je construis du sens (réagir à un texte)

CCI - ITEM 11



Composante 2 : Je porte un jugement critique

- Évaluer la crédibilité des sources et la véracité de l'information dans les textes courants
- Mettre en contexte une œuvre, un texte, un extrait, un propos ou une problématique

Exemple d'item:

Si un ami désirait s'informer sur le phénomène de la consommation, lui proposerais-tu de lire ce texte ? Quelle que soit ta réponse, tu dois la justifier à l'aide de deux critères. Donne un exemple pour chacun de tes critères.

Composante 2 : Je porte un jugement critique

5 POINTS

- L'élève justifie très bien sa réponse à l'aide de deux critères bien illustrés qui touchent à la qualité de l'écriture, à la qualité de l'information, à la présentation visuelle ou au message transmis.

Exemple : Oui, je lui proposerais ce texte parce qu'il est facile de s'y retrouver grâce aux intertitres, mais aussi parce que l'auteure semble s'y connaître sur le sujet : elle apporte beaucoup d'arguments et son texte a été publié dans une revue.

Composante 2 : Je porte un jugement critique

4 POINTS

- L'élève justifie bien sa réponse, à l'aide de deux critères dont l'un est plus ou moins bien illustré, critères touchant toujours à la qualité de l'écriture, à la qualité de l'information, à la présentation visuelle ou au message transmis.

Exemple : *Non, je ne lui proposerais pas le texte, d'abord parce qu'il est mal divisé, mais aussi parce que le vocabulaire est trop compliqué, ce qui fait que j'ai eu de la difficulté à tout comprendre.*

Composante 2 : Je porte un jugement critique

3 POINTS

- L'élève justifie sa réponse en évoquant un seul critère bien illustré ou deux plus ou moins bien illustrés, mais qui touchent toujours à la qualité de l'écriture, à la qualité de l'information, à la présentation visuelle ou au message transmis.

Exemple : *Oui, je lui proposerais ce texte parce que les informations qu'on y retrouve sont très pertinentes pour comprendre le sujet.*

Composante 2 : Je porte un jugement critique

2 POINTS

- L'élève donne un ou deux critères qui justifient peu son opinion OU qui relèvent d'impressions personnelles, non pas d'éléments du texte ou qui ne sont pas illustrés.

Exemple: *Oui, je proposerais ce texte, car l'auteure semble intelligente. Non, je ne proposerais pas ce texte, car il est vraiment plate.*

Composante 2 : Je porte un jugement critique

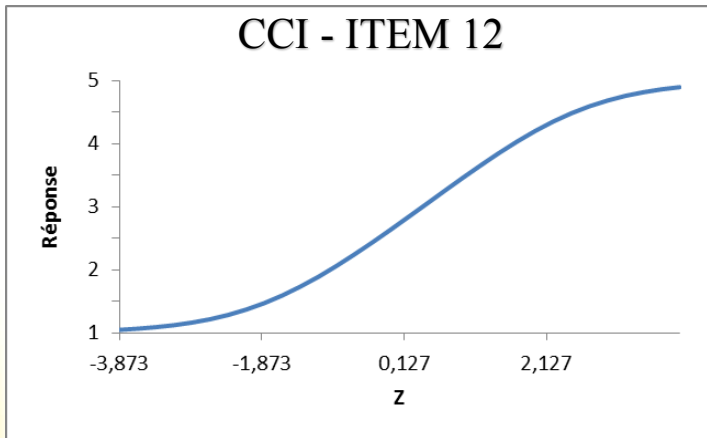
1 POINT

- L'élève ne fait que donner son opinion, sans la justifier à l'aide de critères.

Je porte un jugement critique

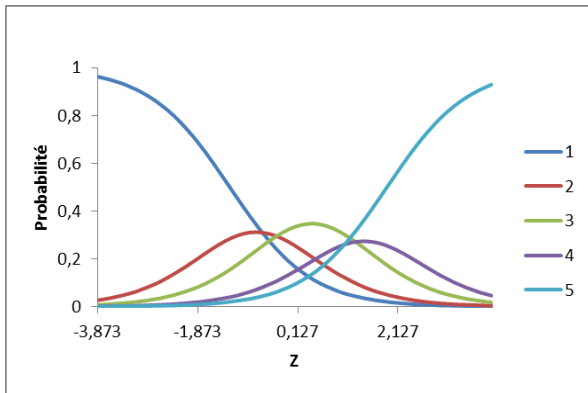
Score	Fréquence	%
1	140	23,3
2	132	21,9
3	163	27,1
4	88	14,6
5	79	13,1
Total	602	100,0

Je porte un jugement critique



Je porte un jugement critique

CCO - ITEM 12



Composante 3 : Je réfléchis à ma pratique de lecteur

- Se donner une démarche adaptée à la situation de communication
- Évaluer l'efficacité de ses stratégies de lecture et d'appréciation

Exemple d'item:

Explique comment tu t'y es pris pour prédire l'idée générale.

Composante 3 : Je réfléchis à ma pratique de lecteur

- Se donner une démarche adaptée à la situation de communication
- Évaluer l'efficacité de ses stratégies de lecture et d'appréciation

Exemple d'item:

Explique comment tu t'y es pris pour prédire l'idée générale.

Composante 3 : Je réfléchis à ma pratique de lecteur

5 POINTS

L'élève nomme une ou des stratégies pertinentes et explicite comment cette ou ces stratégies lui ont permis de prédire l'idée générale.

Exemple par rapport au titre : *J'ai tenu compte du titre, qui évoque les adolescents, la publicité et la surconsommation. Surconsommer signifie « trop consommer ».*

Composante 3 : Je réfléchis à ma pratique de lecteur

4 POINTS

- L'élève nomme une ou des stratégies pertinentes et explicite peu comment cette ou ces stratégies lui ont permis de prédire l'idée générale.

Exemple : *J'ai tenu compte du chapeau du texte (i.e. le 2^e élément le plus important après le titre), qui parlait du sujet que j'ai nommé.*

Composante 3 : Je réfléchis à ma pratique de lecteur

3 POINTS

- L'élève nomme une ou des stratégies pertinentes, mais n'explique pas comment cette ou ces stratégies lui ont permis de prédire l'idée générale

Exemple : *J'ai tenu compte du titre et/ou du chapeau et/ou des intertitres et/ou des premières phrases de chaque paragraphe et/ou des images.*

Composante 3 : Je réfléchis à ma pratique de lecteur

2 POINTS

- L'élève nomme une ou des stratégies non pertinentes dans le cas du texte à l'étude, celui-ci présentant plusieurs indices de prédiction manifestes.

Exemple : *J'ai parcouru le texte OU J'ai lu jusqu'à ce que le temps soit écoulé.*

Composante 3 : Je réfléchis à ma pratique de lecteur

1 POINT

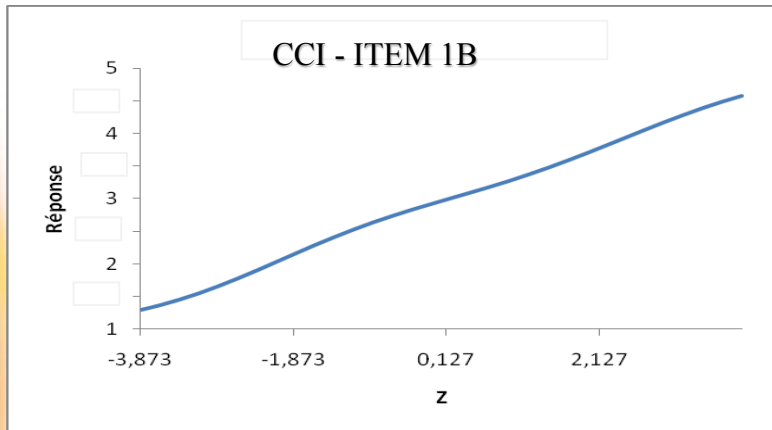
- L'élève nomme une ou des stratégies non pertinentes dans le cas du texte à l'étude, celui-ci présentant plusieurs indices de prédiction manifestes.

Exemple : L'élève ne nomme pas de stratégie.

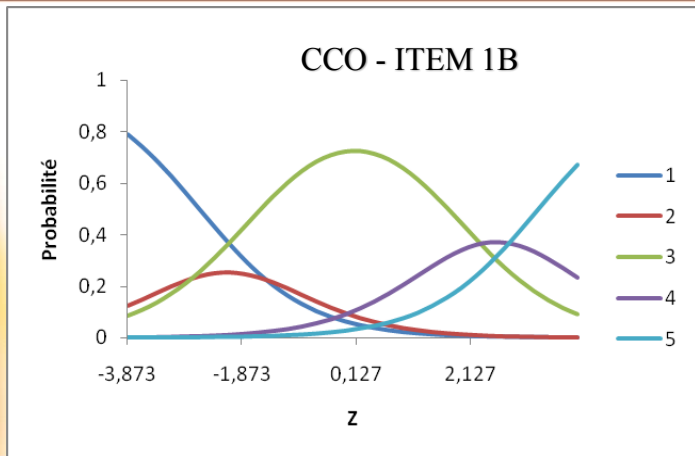
Je réfléchis à ma pratique de lecteur

Point	Fréquence	Pourcentage
1	55	9,1
2	65	10,8
3	386	64,1
4	70	11,6
5	26	4,3
Total	602	100,0

Je réfléchis à ma pratique de lecteur



Je réfléchis à ma pratique de lecteur



Composante 4 : Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

- S'appuyer sur ses connaissances relatives aux régularités de la langue et des textes pour en acquérir de nouvelles

Exemple d'item:

À la lecture de ce texte, qu'as-tu appris concernant la publicité et son influence sur les jeunes?

Composante 4 : Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

3 POINTS

- L'élève est capable de nommer de façon juste ce qu'il a appris, sans toutefois retranscrire un ou des passages textuellement.

Exemple : *J'ai appris que les compagnies n'ont pas le droit de faire de la publicité directement auprès des enfants.*

Composante 4 : Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

2 POINTS

- L'élève est capable de nommer de façon juste ce qu'il a appris. Cependant, il ne fait que retranscrire un ou des passages textuellement.

Exemple: *J'ai appris que cumulés, ces achats à long terme décevants peuvent avoir des effets néfastes sur les jeunes.*

Composante 4 : Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

1 POINT

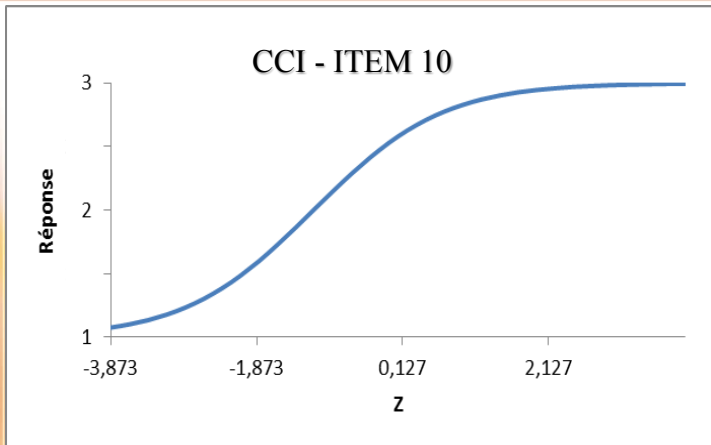
- L'élève donne une réponse qui n'entretient pas de lien avec le texte (indice d'une mauvaise compréhension) OU il fait un commentaire impertinent.

Exemple : *J'ai appris que les enfants consomment dans leur berceau.* OU *Je n'ai rien appris, c'était ennuyeux.*

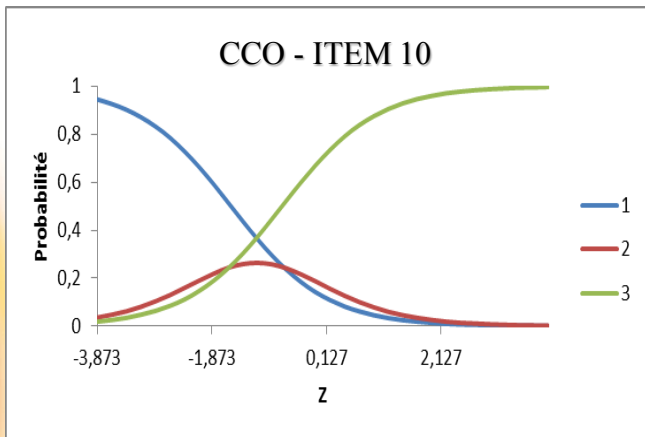
Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

Score	Fréquence	%
1	113	18,8
2	99	16,4
3	390	64,8
Total	602	100,0

Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture



Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture



Conclusion

- Difficulté de bien définir les concepts
- Difficulté à bien définir les critères d'évaluation
- Utilité des modèles de mesure pour rendre les évaluations **plus objectives**