Présentation du modèle de Rasch

par Helmut Vorkauf, IAE: Institut de recherche en éducation et examens, Université de Berne

Présentation du modèle

La proposition fondamentale du modèle proposé par le mathématicien **Georg Rasch** (1960) est extrêmement simple : quand une personne rencontre une question de test («item»), la *chance* d'une réponse correcte est le produit de deux paramètres :

f_i, le paramètre de facilité de l'item i

a_j, le paramètre d'aptitude de la personne j.

Ainsi donc : chance d'une réponse correcte = $f_i \times a_j$.

Plus l'aptitude et la facilité sont grandes, plus grande est la chance d'une réponse correcte. Si, par exemple, a = 2 et f = 3, la chance devient 2×3 = 6, ce qui signifie qu'on pourrait faire un pari avec une chance de 6 contre 1 sur une réponse correcte de cette personne.

On a coutume de parler plutôt des *probabilités* : une probabilité p est définie $$p = \frac{chance}{1 + chance}$$

Autre formulation du modèle, utilisant cette définition :

$$p = \frac{a_j f_i}{1 + a_j f_i} \quad \text{ou aussi :} \quad p = \frac{1}{1 + \frac{1}{a_j f_i}}$$

Dans la littérature, on trouve beaucoup d'autres lettres latines et grecques pour les paramètres, de même que pour les logarithmes de a et de f, ce qui ne simplifie pas la lecture pour le débutant. Si α et φ sont les logarithmes de a et f, la formule pour p devient

$$p = \frac{e^{\alpha+\varphi}}{1 + e^{\alpha+\varphi}} \quad \text{ou aussi} \quad p = 1 / (1 + 1 / e^{\alpha+\varphi})$$

D'autres préfèrent un paramètre de difficulté d au lieu de la facilité f. Si on définit la difficulté d = 1/f, la formule devient

$$p = \frac{a / d}{1 + a / d}$$

Avec un peu d'algèbre, on peut toujours passer, par transformation, d'une formulation à l'autre.

Qu'est-ce que cela implique pour les propriétés statistiques des items ? L'analyse d'items que nous connaissons (appelée ici *analyse classique*) reposait sur le postulat d'une discrimination positive : il fallait que la réponse correcte soit plus attractive pour les personnes les plus capables. Par exemple, plus un élève sait d'algèbre, plus grande est la probabilité qu'il trouvera la réponse correcte à tel item spécifique.

Si, avec une aptitude croissante, la proportion des réponses correctes augmente, l'item a une discrimination positive, quelle que soit la fonction mathématique qui lie la proportion à l'aptitude. Il suffit d'avoir une augmentation monotone. Plus la pente est forte, plus l'item est discriminant. L'idéal, dans la représentation graphique, serait en «marche d'escalier» : jusqu'à un certain degré d'aptitude, les élèves répondent tout à fait faux ; depuis tel degré, ils répondent tout à fait juste.

En contraste, le modèle de Rasch demande que l'augmentation de la proportion des réponses correctes suive la *fonction logistique*

$$p = 1/(1 + 1/(a_j e_i))$$

qui apparaît comme une fonction sigmoïde avec une croissance d'abord accélérée et puis désaccélérée. Les fonctions des items de différentes facilités sont parallèles l'une à l'autre, comme les trois items (a), (b) et (c), à la figure 1a.

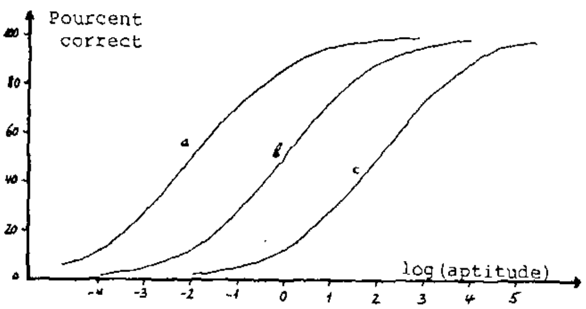


Fig. 1a: Fonctions logistiques de trois items de différentes facilités.

D'autres formes de fonction sont possibles, et on les trouve dans l'analyse des tests. A la figure 1b, nous avons trois types de fonction qui ne suivent pas le modèle de Rasch : l'item (a) ne commence pas à zéro pour cent de réussite, probablement à cause de réponses correctes au hasard; l'item (c) a une discrimination négative; et l'item (b) a une croissance trop rapide. Selon la théorie classique, l'item (b) serait idéal ; selon la théorie de Rasch, il faudrait l'enlever parce qu'il ne suit pas la fonction citée. Il est trop discriminant, constatation absurde dans la théorie classique.

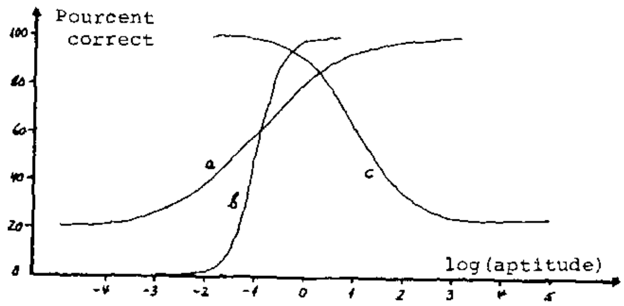


Fig. 1b: Fonctions non logistiques de trois items mal adaptés au modèle de Rasch.

Si, dans la réalité empirique, des items se trouvent avoir des *courbes caractéristiques* qui suivent la fonction logistique, on peut tirer profit, dans les applications, de quelques propriétés particulièrement remarquables du modèle. Nous reviendrons sur ce point dans une autre partie de cet article.

Auparavant, il faut se poser la question de savoir si le modèle représente la réalité empirique. On a trouvé bien des tests dont la plupart des items ont réellement des courbes caractéristiques logistiques. Notre propre expérience est limitée à des tests de connaissance et de compréhension en médecine, où l'on constate une régularité remarquable. La plupart des items y ont vraiment des courbes caractéristiques qui sont en accord avec le modèle logistique de Rasch.

La figure 2a présente un exemple représentatif d'un item assez bien adapté, la figure 2b est un exemple d'un item mal adapté au modèle (pris dans un test d'ophtalmologie).

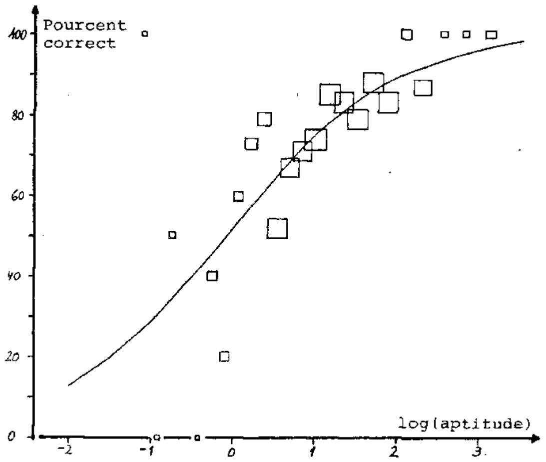


Fig. 2a:
Item bien adapté.

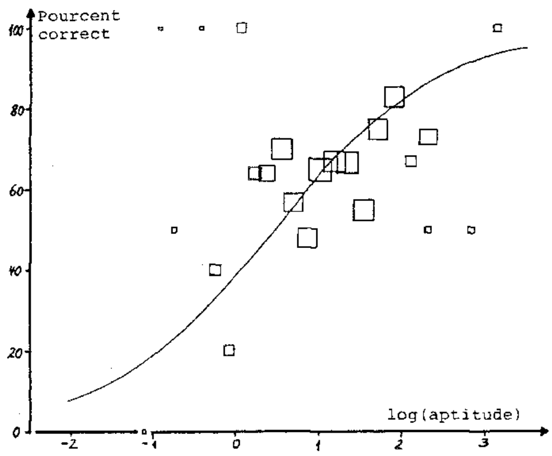


Fig. 2b:
Item mal adapté.

Si les items ont des courbes caractéristiques de forme logistique, on peut constater la validité du modèle de Rasch. Est-ce là une simple constatation intéressante ou peut-on en tirer un profit pratique ?

Indépendance de l'échantillon

C'est la qualité supérieure du modèle. Revenons à l'équation de base

$$p_1 = \frac{a_j f_i}{1 + a_j f_i} \qquad \text{(probabilité d'une réponse correcte)}$$

$$p_0 = \frac{1}{1 + a_j f_i} \qquad \text{(probabilité d'une réponse fausse)}$$

Quelle est la probabilité p_{11} de deux réponses correctes à deux items de facilité f_1 et f_2 pour une personne d'aptitude a ?

On multiplie entre elles les probabilités d'une réponse correcte à chacun des deux items

$P_{11} = (af_1/(1 + af_1)) \times (af_2/(1 + af_2))$

Par analogie, on arrive aux probabilités d'une réponse correcte au premier item et non au deuxième P_{10}, d'une réponse correcte au deuxième et non au premier p_{01}, et de deux réponses fausses p_{00} :

$P_{10} = (af_1/(1 + af_1)) \times (1/(1 + af_2))$

$P_{01} = (1/(1 + af_1)) \times (af_2/(1 + af_2))$

$P_{00} = (1/(1 + af_1)) \times (1/(1 + af_2))$

La probabilité d'avoir une seule réponse correcte, que ce soit au premier item ou au deuxième, devient

$P_{10} + P_{01} = a\,(f_1 + f_2) / ((1 + af_1)(1 + af_2))$

Finalement, la probabilité conditionnelle de répondre correctement au premier item sous condition de répondre correctement à l'un ou l'autre, mais non à tous les deux, résulte en

$$\frac{p_{10}}{p_{10} + p_{01}} = \frac{af_1 / ((1 + af_1)(1 + af_2))}{a(f_1 + f_2) / ((1 + af_1)(1 + af_2))} = \frac{f_1}{f_1 + f_2}$$

Cette formule élimine le paramètre d'aptitude a : la comparaison de la facilité des deux items est indépendante des personnes qui répondent aux items si les personnes sont d'aptitude homogène. C'est ce que Rasch appelle *indépendance stochastique locale***.** Autrement dit, la corrélation entre deux items est zéro quand toutes les personnes sont d'aptitude égale.

Après tant d'algèbre, retournons à la réalité d'un test administré à des personnes en chair et en os. Si l'algèbre est correcte, la facilité des items d'un test doit être pareille pour deux groupes d'étudiants, même si les deux groupes sont des échantillons très différents. Dans un test d'ophtalmologie de 40 items, on a éliminé 12 items mal adaptés comme l'item de la figure 2b. Utilisant les 28 items restants, on a fait deux groupes d'étudiants, ceux qui ont un score de 0 à 18 et ceux qui ont de 19 à 28 points.

Evidemment, la facilité des items évaluée selon la théorie classique (pourcent de réponses correctes) est très différente pour les deux groupes (fig. 3a). La théorie classique ne connaît pas l'indépendance de l'échantillon.

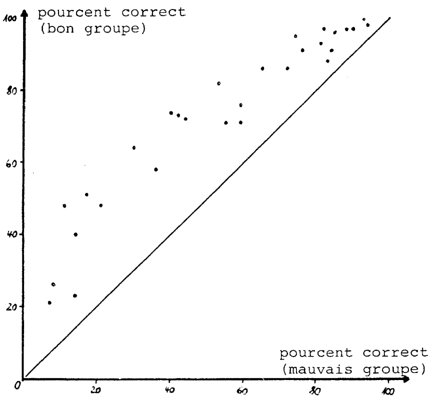


Fig 3a: Comparaison des facilités classiques en deux groupes d'aptitude différente

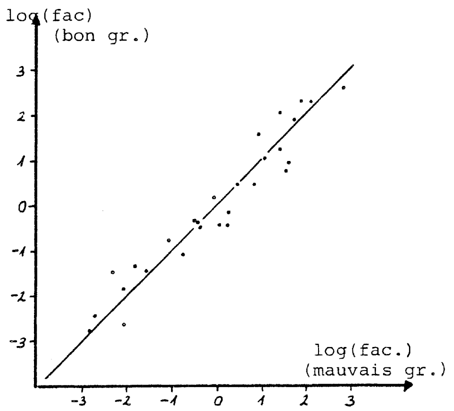


Fig 3b: Comparaison des facilités selon Rasch en deux groupes d'aptitude différente

Voici maintenant la comparaison des deux groupes d'étudiants quant à l'estimation indépendante de la facilité suivant Rasch (fig. 3b). Il n'y a presque plus de différences entre les deux estimations indépendantes, les deux estimations de la facilité sont, dans des limites statistiques, égales grâce à l'indépendance de l'échantillon de personnes.

Dans le modèle, les paramètres d'aptitude et de facilité sont en rapport de multiplication, c'est-à-dire de symétrie. De même que la facilité d'un item est estimée indépendamment de l'échantillon des personnes, de même on peut estimer l'aptitude d'une personne d'une manière indépendante de l'échantillon des items. Quand on sait les efforts qu'exigé la construction des tests parallèles dans la théorie classique, ce caractère du modèle de Rasch constitue une promesse presque incroyable.

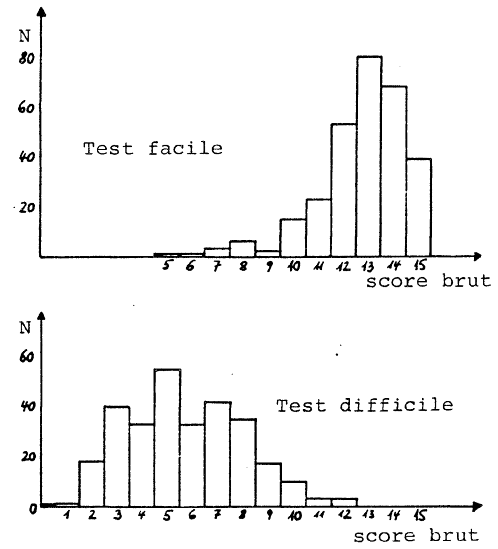


Fig 4a: Distribution du score brut des deux tests de difficultédifférente

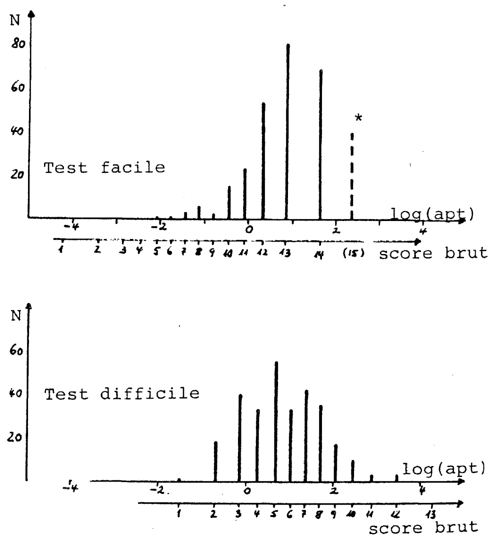


Fig 4b: Distribution du logarithme du paramètre d'aptitude de deux tests de difficulté différente (*: on ne peut pas estimer l'aptitude associée à un score parfait)

Reprenons le test d'ophtalmologie. On sépare deux moitiés du test, une partie composée des items faciles, l'autre des items difficiles. La distribution des étudiants en score brut (nombre de réponses correctes) des deux sous-tests est extrêmement différente (fig. 4a), mais dans les deux cas, l'aptitude estimée des étudiants est comparable, ce qui est démontré dans la figure 4b.

POSSIBILITÉS DANS L'APPLICATION

Toutes ces caractéristiques sont intéressantes théoriquement. Mais à quoi servent-elles dans la pratique ?

Economie

Dans la construction d'un test, une des choses qui coûtent très cher, c'est de se procurer un échantillon représentatif de personnes pour la calibration du test.

L'indépendance de l'échantillon introduit une grande liberté : n'importe quel groupe de personnes est valable, pourvu qu'il fasse partie de la population. (L'indépendance de l'échantillon ne me donne pas la liberté d'essayer un test de chirurgie sur des enfants de 7 ans.)

Une fois qu'un pool (banque) d'items est calibré, tout groupe d'items tiré de ce pool se trouve aussi calibré. Pour chaque nouvelle forme «parallèle» de test construit d'après la théorie classique, il fallait une calibration avec un échantillon représentatif de personnes. Dans le modèle de Rasch, il suffit d'un petit calcul.

Concaténation de deux tests

Le problème de l'équivalence des tests, dans le modèle de Rasch, a perdu ses dangers. Pour trouver une échelle commune à deux ou plusieurs tests, il suffit que ceux-ci aient quelques items en commun. Lorsqu'on change un curriculum d'une année à l'autre, il suffit qu'il reste un noyau commun d'items pour garantir le même niveau d'évaluation.

Le problème de la perte de discrimination au niveau de la maîtrise

Dans des tests centrés sur des critères, on a souvent des problèmes de variance qui disparaît. Au pré-test, les élèves ont peu de réponses bonnes; au post-test, ils en ont trop. Les analyses classiques ne donnent plus de résultats raisonnables.

Avec Rasch, on évite de donner des questions qui sont trop difficiles (au pré-test) ou trop faciles (au

post-test). Dans le pool des items, on peut choisir des items dont la facilité soit appropriée au niveau des élèves et ensuite on peut estimer la probabilité que ceux-ci répondent correctement aux items qui ne sont pas encore (pré-test) ou plus (post-test) appropriés.

Admettons que les facilités du pool se situent entre 0.05 et 20.0, avec une moyenne de 1.0. Dans le pré-test, on a choisi des items faciles entre 17.0 et 20.0. Un élève a un score qui représente une aptitude de 7.0. Dans un test normal composé d'items de facilité 1.0, il a une probabilité de trouver les bonnes réponses de $p = \frac{a_j f_i}{1 + a_j f_i} = \frac{7 \times 1}{1 + (7 \times 1)} = 0.88$

Dans un échantillon représentatif de 100 items, son score le plus probable serait 88, c'est ce que nous appelons la *valeur standard***.**

Si la maîtrise est définie comme 95% (95% du pool), on ne pourrait constituer l'examen que d'items qui soient encore suffisamment difficiles pour avoir assez de discrimination. Mais même un maître ne répondra pas à 95% de ces questions ; cependant si l'estimation donne une probabilité de 95% du pool, on peut déclarer avoir atteint le niveau de la maîtrise.

APPENDICE A

Estimation des paramètres de facilité

Cette estimation se fait normalement grâce à la solution par itération des équations suivantes :

$$n_{i.} = \sum_{j=1}^{k-1} \frac{n_{.j} a_j f_i}{1 + a_j f_i} \qquad\qquad j = \sum_{i=1}^{k} \frac{a_j f_i}{1 + a_j f_i}$$

ou $n_{i.}$ = nombre de personnes avec la réponse correcte à l'item i;

$n_{.j}$ = nombre de personnes avec j réponses correctes;

k = nombre d'items;

a_j = aptitude associée à j réponses correctes;

f_i = facilité de l'item;

$\sum$ représente le signe de sommation.

Cette approximation itérative ne peut être effectuée que sur un ordinateur, mais plusieurs programmes rapides sont à disposition (v. App. F).

Une méthode plus simple pour le calcul manuel se trouve chez **Fischer** sous le titre de «méthode explicite».

Comme nous l'avons vu, la probabilité conditionnelle d'une réponse correcte à l'item i, sous condition d'une seule réponse correcte sur deux items i et j, est $\frac{p_{10}}{p_{10} + p_{01}} = \frac{f_i}{f_i f_j}$

Si n_{ij} est le nombre de personnes avec une bonne réponse à l'item i et une mauvaise réponse à l'item j, et si n_{ji} est le nombre de personnes avec une réponse correcte à l'item j et non à l'item i, le nombre de personnes avec un score de 1 est égal à $n_{ij} + n_{ji}$. Le nombre total de personnes est égal à N, et la proportion de réponses correctes est $(n_{ij} + n_{ji})/N$.

Cette proportion peut servir d'estimation de $(P_{10} + P_{01})$.

Par analogie, n_{ij}/N est une estimation de p_{10} et n_{ji}/N est une estimation de p_{01}.

L'insertion de ces estimations dans la formule pour la probabilité conditionnelle aboutit à

$$\frac{n_{ij}}{(n_{ij} + n_{ji}) / N} = \frac{f_i}{f_i + f_j}$$

Si on échange i et j $$\frac{n_{ji} / N}{(n_{ij} + n_{ji}) / N} = \frac{f_j}{f_i + f_j}$$

La division de ces deux dernières équations donne $$\frac{n_{ij}}{n_{ji}} = \frac{f_i}{f_j}$$

La généralisation pour k items, au lieu de deux seulement, nous donne l'équation

$$\prod_{\substack{j=1\\ j\neq i}}^{k} \frac{n_{ij}}{n_{ji}} = \frac{f_i^{\,k-1}}{\prod_{\substack{j=1\\ j\neq i}}^{k} f_j}$$ (Π représente le signe de multiplication).

La multiplication avec f_i/f_j se simplifie en $$\prod_{\substack{j=1\\ j\neq i}}^{k} \frac{n_{ij}}{n_{ji}} = \frac{f_i^{\,k}}{\prod_{j=1}^{j=k} f_j}$$

En définissant le produit de tous les f_j: $$\prod_{j=1}^{k} f_j = 1$$ on obtient $$\prod_{\substack{j=1\\ j\neq i}}^{k} \frac{n_{ij}}{n_{ji}} = f_j^{\,k}$$

ou $$f_j = \{\prod_{\substack{j=1 \\ j \neq i}}^{k} \frac{n_{ij}}{n_{ji}}\}^{1/k}$$

Voici un exemple pratique de cette méthode explicite de Fischer : on a construit une matrice (n_{ij}) de fréquences, où l'élément dans la ligne i et la colonne j représente le nombre de personnes qui ont bien répondu à l'item i et mal à l'item j.

	1	2	3	4	5
1	-	11	4	1	2
2	32	-	11	3	7
3	43	29	-	6	3
4	58	39	24	-	8
5	62	46	24	11	-

Dans cet exemple, il y a 6 personnes qui ont répondu correctement à l'item 3, mais non àl'item 4.

La facilité du premier item, d'après la dernière formule, est

$$f_1 = \{\prod_{\substack{j=1 \\ j \neq i}}^{5} \frac{n_{ij}}{n_{ji}}\}^{1/5} = \{\frac{11 \times 4 \times 1 \times 2}{32 \times 43 \times 58 \times 62}\}^{1/5} = 0.1122$$

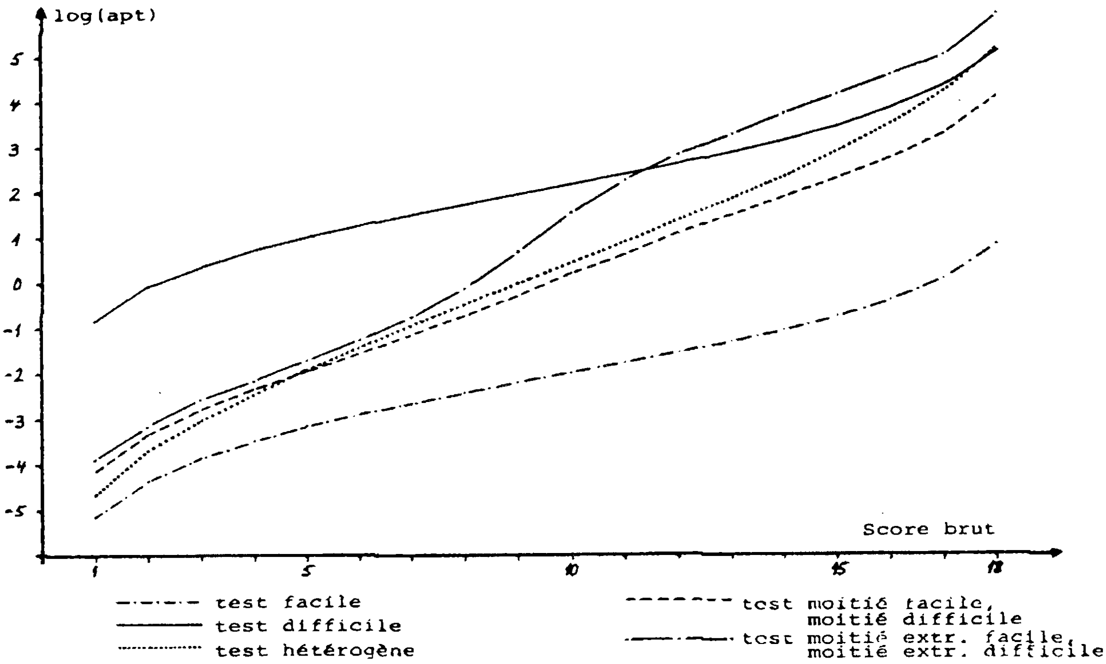
Les résultats pour tous les cinq items : f1=0.1122, $f_{2=}$ 0.4190, f_3=1.0246, f_4=4.6583, f_5=4.6782

(Malheureusement cette méthode simple n'est plus applicable quand une ou plusieurs des fréquences n_{ij} deviennent zéro.)

APPENDICE B :

Estimation des paramètres d'aptitude

L'estimation des paramètres d'aptitude normalement n'est faite qu'après que les paramètres de facilité



des items sont connus. Une application pratique de l'estimation des paramètres d'aptitude seulement, c'est la construction d'une table de normes d'un test qui consiste en un échantillon d'items tirés d'un pool d'items calibrés. La solution de ce problème est une solution de **maximum likelihood,** c'est-à-dire que les paramètres sont estimés de telle manière que la probabilité de l'échantillon des données soit maximisée, sous condition de paramètres de facilité connus.

Cette estimation **maximum likelihood** exige la solution des équations suivantes pour les a_j :

$$j = \sum_{i=1}^{k} \frac{a_j f_i}{1 + a_j f_i}$$

i = 1,2, 3,..., k;

j = 1,2, 3, ..., (k-1);

k = nombre d'items;

j = score brut (nombre de réponses correctes);

a_j = aptitude associée au score j;

f_i = facilité de l'item i.

On ne peut pas résoudre directement ce système de (k-1) équations, mais on peut s'approcher des solutions en quelques itérations en employant la méthode de Newton-Raphson:

$a_{n+1} = a_n - \dfrac{g(a_n)}{g'(a_n)}$, où n désigne la n^e itération, a_n est le paramètre d'aptitude dans la n^e itération,

$$g(a) = j - \sum_{i=1}^{k} \frac{a_j f_i}{1 + a_j f_i} \qquad g'(a) = -\sum_{i=1}^{k} \frac{a_j f_i}{(1 + a_j f_i)^2}$$

On peut finir le cycle d'itérations quand la valeur absolue des différences des a_j obtenues dans deux cycles consécutifs devient assez petite, par exemple plus petite que 10^{-6}.

Il est peut-être suggestif de comparer les aptitudes estimées de différents tests. Quelle est l'aptitude associée aux scores de 1 à 19 des tests suivants : un test facile, composé de 19 items de log(f) (logarithme de facilité) de 2.000 ; un test difficile : 19 items de log(f) = -2.000 ; un test hétérogène, où les 19 items ont des log(f) de -4.500, -4.000, -3.500, ..., 0.000, .., +4.000, +4.500; un test extrêmement hétérogène, où 9 items sont très difficiles (log(f) = -3.900), et où 10 autres items sont faciles (log(f) = +1.900).

Les courbes qui représentent l'association de l'aptitude au score brut se trouvent à la figure B1. On voit que le test facile fait une discrimination au niveau bas de l'échelle de l'aptitude, le test difficile au niveau élevé. Le test hétérogène peut mesurer dans une partie plus grande de l'échelle. Le mesurage avec les tests homogènes (faciles et difficiles) est limité à un niveau d'aptitude spécifique, mais ce mesurage limité est plus fin : les courbes des tests homogènes ont peu de pente.

Finalement, le test extrêmement hétérogène, qui est en fait un test composé de deux tests de facilité différente, a une courbe qui consiste en deux courbes liées.

APPENDICE C :

Tests de la validité du modèle

La décision quant à l'appartenance d'un item spécifique à une échelle Rasch repose sur les propriétés de sa courbe caractéristique. Si cette courbe est bien représentée par la fonction logistique, la facilité de l'item doit être égale pour des sous-échantillons de personnes. C'est cette indépendance de l'échantillon qui est investiguée par la plupart des tests proposés.

Supposons que, dans un test de 7 items, 368 étudiants ont obtenu des scores de 1 à 6 réponses correctes, et que le nombre de réponses correctes, reparti par groupe d’étudiants avec un score de 1 à 6, soit le suivant :

	Items							nombre de
Score	1	2	3	4	5	6	7	personnes
1	0	0	1	5	0	1	0	7
2	1	6	6	5	7	3	4	16
3	21	15	10	21	11	3	3	28
4	53	40	43	43	48	20	13	65
5	110	106	106	89	92	65	22	118
6	133	129	130	124	125	116	47	134
Total	318	296	296	287	283	208	89	368

Par exemple, des 134 étudiants avec un score de 6, 133 ont donné la réponse correcte au premier item.

L'estimation des paramètres de facilité, basée sur tous les 368 étudiants, converge sur les facilités (logarithmes) :

Item	1	2	3	4	5	6	7
log(facilité)	1.187	0.651	0.651	0.465	0.387	-0.808	-2.532

Pour une investigation de l'indépendance de l'échantillon, on pourrait estimer les paramètres encore une fois, cette fois en utilisant seulement les données des étudiants avec des scores de 1, 2, et 3.

La matrice des données prend la forme :

Score	1	2	3	4	5	6	7	personne
1	0	0	1	5	0	1	0	7
2	1	6	6	5	7	3	4	16
3	21	15	10	21	11	3	3	28
Total	22	21	17	31	18	7	7	51

L'estimation des paramètres de facilité, basée sur ce sous-échantillon seulement, donne les résultats :

Item	1	2	3	4	5	6	7
log(facilité)	0.483	0.397	0.039	1.254	0.131	-1.152	-1.152

En se basant sur les données des étudiants avec des scores de 4 et 5, et sur ceux avec un score de 6, on arrive à deux autres estimations indépendantes, qui sont soumises à une comparaison.

	Estimation du logarithme des paramètres de facilité, basée sur les données des sous-échantillons		
Item	1 à 3	4 à 5	6
1	.483	1.359	2.316
2	.397	.614	.676
3	.039	.722	.907
4	1.254	.176	-.056
5	.131	.414	.057
6	-1.152	-.970	-.711
7	-1.152	-2.316	-3.190

On reconnaît que l'estimation du paramètre de facilité des items 1,4, et 7 dépend de l'échantillon : la facilité de l'item 1 augmente avec l'aptitude (autrement dit : il est trop discriminatif), celle des items 4 et 7 est trop grande pour les étudiants faibles (ces items sont peu discriminants).

Il va de soi qu'il est permis de former des sous-échantillons d'autres manières, par exemple par sexe ou par langue maternelle, pour découvrir des inadaptations au modèle. Dans les exemples mentionnés, des inadaptations peuvent être interprétées dans le sens d'un biais en faveur de certains groupes.

APPENDICE D :

L'usage d'une banque d'items et la concaténation de deux tests.

Supposons qu'on a une banque d'items, c'est-à-dire un pool d'items dont on connaît les paramètres de facilité. Dans ce pool, on a les items 1, 2, et 3 avec les facilités

0.14 0.37 1.00

On assemble un test de disons cinq items, composé de ces trois items tirés du pool et deux nouveaux items 4 et 5. L'analyse donne une estimation primaire des 5 facilités

0.22 0.61 1.65 1.65 2.72

Le produit de ces cinq paramètres initiaux est, comme d'habitude, égal à 1. La facilité des trois items connus est surestimée (par exemple 0.22 au lieu de 0.14 pour l'item 1), mais cela n'est pas gênant, car la norme initiale est complètement arbitraire. Il suffit d'une correction des estimations primaires par une constance c

$c = ((0.14 \times 0.37 \times 1.00) / (0.22 \times 0.61 \times 1.65))^{1/3} = 0.616$, ce qui nous donne les facilités définitives

0.14 0.38 1.02 1.02 1.68

On utilise ces facilités calibrées pour estimer les aptitudes (cf. app. B).

Ainsi, le nouveau test est calibré à la norme de la banque d'items, et la banque d'Items a gagné deux nouveaux items qui sont utilisables comme des items calibrés dans de futurs échantillons du pool.

La même technique peut être utilisée pour la concaténation de deux tests où les deux tests ont quelques items en commun. Dans le cas extrême, il suffirait d'un seul item commun.

APPENDICE E : **Autres modèles**

Le modèle de Rasch n'est pas le seul modèle probabilistique. On peut construire des modèles plus élaborés, par exemple en introduisant un paramètre de discrimination d'un item ou/et un paramètre de l'item qui représente la tendance à provoquer la réponse correcte au hasard. Dans un tel modèle, les items avec des courbes caractéristiques comme a et b de la fig. 1b peuvent être retenus. Mais, à notre avis, on paie cher pour la rétention de quelques items : l'indépendance de l'échantillon est sacrifiée, qui est pourtant la raison la plus importante pour adopter un tel modèle. Il semble plus sage de se borner aux items de discrimination uniforme pour tirer profit de l'indépendance.

APPENDICE F : **Programmes d'ordinateur**

La solution des systèmes d'équations complexes par itérations n'est presque jamais effectuée sans ordinateur. Heureusement, il y a déjà quelques programmes d'ordinateur qui rendent le travail plus facile.

Le programme le plus simple est celui de **Wright and Panchapakesan** (1969) ; **Fischer** (1974) a aussi publié divers algorithmes. Dans **Fischer** (1974), on trouve aussi quelques programmes pour des cas spéciaux, par exemple pour le cas des items à réaction polychotome, où les réactions ne sont pas limitées à la dichotomie de réponses correctes et fausses.

Hambleton et Cook (1977) citent les sources où on peut obtenir les programmes pour des modèles à plusieurs paramètres, dont le modèle de Rasch n'est qu'un cas spécial. Tous ces programmes ont le désavantage qu'ils coûtent cher en temps d'ordinateur et que les algorithmes ne convergent pas toujours sur une solution.

Pour un débutant, le modèle simple de Rasch et un programme comme celui de Wright et Panchapakesan ou un des programmes de Fischer sont probablement plus recommandables.

(On peut demander l'assistance de l'auteur de cet article, qui dispose de copies et modifications de ces programmes.)

Références:

FISCHER, G.: *Einführung in die psychologische Testtheorie,* Bern, Huber 1974.

HAMBLETON, R. K. et COOK, L L: *Latent Trait Models and their Use in the Analysîs of Educational Test Data,* J. Educ. Measmt, 1977, 14, 75-96.

RASCH, G.: *Probabilistic Models for some Intelligence and Attainment Tests,* Copenhagen, Danmarks Paedagogiske Institut, 1960.

WRIGHT, B. D. et PANCHAPAKESAN, N.: *A Procedure for Sample-free Item Analysis,* Educ. Psychol. Measmt, 1969, 29, 23-48.