

CRITIQUE DES METHODES D'APPLICATION,
DE CORRECTION ET DE COTATION DES QUESTIONS
A CHOIX MULTIPLE

par

Dieudonné LECLERCQ
(Université de Liège)

INTRODUCTION

Une certaine rigidité dans l'évaluation des questionnaires à choix multiple peut exercer une influence éducative défavorable. Par exemple, si s'abstenir de répondre est aussi durement sanctionné que fournir une réponse incorrecte, l'élève a avantage à répondre aveuglément plutôt que de reconnaître honnêtement une ignorance. En outre, dans les pays anglo-saxons, l'élève est presque toujours autorisé à choisir dans un ensemble un certain nombre de questions et peut donc obtenir le maximum des points sans avoir répondu à tout ; ce n'est point l'habitude chez nous.

Or, à notre époque, il est plus pertinent de reconnaître son ignorance que d'afficher de fausses certitudes. Comment jugerait-on un médecin généraliste qui refuserait d'envoyer un patient chez un spécialiste sous prétexte qu'il s'agirait là d'un aveu d'incompétence ? Et le mécanicien qui laisserait croire au pilote que son avion est en état de vol, alors qu'il n'en saurait rien ? Pareille attitude est pourtant tolérée, voire encouragée ou imposée par l'emploi de mauvaises procédures de construction, d'application, de correction et de cotation des épreuves scolaires. Ces quatre aspects de l'examen sont étroitement liés : ils dépendent du modèle qui a servi de référence. Il importe de rendre la mesure « moralement plus juste et scientifiquement plus exacte »¹.

1. G. DE LANDSHEERE, *Évaluation continue et examens, Précis de didactologie* (Bruxelles, Labor, 1971), p. 7.

Récemment, B. CHOPPIN¹ a décrit trois modèles qui ont inspiré les différentes procédures :

Modèle 1 : Si l'élève connaît la bonne solution, il la choisit ; s'il ne la connaît pas, il omet ou il choisit au hasard parmi les solutions proposées.

Modèle 2 : Si l'élève connaît la bonne solution, il la choisit ; sinon il élimine les solutions qu'il sait incorrectes et choisit ensuite au hasard parmi les solutions qui restent.

Modèle 3 : L'élève évalue l'attractivité de chaque question et assigne à chacune une probabilité d'être correcte ; il décide ensuite selon les probabilités qui lui paraissent les plus élevées.

Nous présenterons et nous critiquerons les procédures les plus connues inspirées de ces modèles ; ensuite nous proposerons une formule qui nous paraît originale.

A. PRATIQUES INSPIRÉES DU MODÈLE 1

On demande à l'élève de choisir parmi les solutions proposées ou de les ordonner entre elles. B. DE FINETTI² cite les procédures suivantes :

- « 1 : La mise en ordre de toutes les solutions proposées (en notant 1, 2, 3... r) devant chacune³.
- 2 : La mise en ordre incomplète (où l'on doit ordonner les m premières solutions).
- 3 : Choisir une seule solution (c'est la procédure la plus répandue).
- 4 : Choisir une solution, ou aucune, ce qui donne lieu à trois niveaux de scores : S + (bonnes réponses), S° (omissions), S - (mauvaises réponses).

1. B. CHOPPIN, *An IEA Study of Guessing, a Proposal* (Stockholm, International Association for the Evaluation of Educational Achievement, 1970).

2. B. DE FINETTI, *Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item*, *British Journal of Mathematical and Statistical Psychology*, Vol. 18 (mai 1965), 87-123.

3. Dans nos exemples, r désigne le nombre de solutions proposées.

Ce qu'on peut savoir d'une telle réponse, c'est que le sujet donne à la solution choisie une probabilité entre $1/r$ et 1 (certitude complète) ».

Bien des *corrections for guessing* ont été élaborées. Nous n'en examinerons que trois :

1. *Correction for guessing selon l'attrance théorique*

Cette formule décrite par de nombreux auteurs¹ consiste à donner un point à l'élève par bonne réponse et à retirer $1/(r-1)$ point par mauvaise réponse. Le total des résultats possibles vaut 0, donc l'espérance mathématique vaut également 0. En effet, pour $r = 4$, la bonne réponse rapporte 1 point et chacune des trois mauvaises réponses rapporte $-1/3$ point. Dans une épreuve de 20 questions ($r = 4$), un élève qui répondrait totalement au hasard aurait, à chaque question, une chance sur quatre de choisir la bonne solution. En théorie, 5 réponses sur 20 seraient correctes et les 15 autres incorrectes. Ainsi l'élève obtiendrait $5 (+1) + 15 (-1/3) = 0$.

Cette pénalisation des mauvaises réponses classe les élèves en trois groupes à chaque question :

- Ceux qui ont réussi : +1 point
- Ceux qui ont omis : 0 point
- Ceux qui ont échoué : $-1/r$ point

Nous pensons qu'il est souhaitable de pénaliser les mauvaises réponses et de faire ainsi la distinction d'avec les omissions. Mais la formule utilisée suppose que les solutions sont également attractives... condition qui n'est pratiquement jamais réalisée. En conséquence, cette procédure « ne fait que tenter de compenser le *guessing*, pas de le pénaliser »² ; le désavantage de ces méthodes conventionnelles est d'encourager le *guessing* »³.

1. Par exemple L. POSTMAN, 'Choice Behavior and the Process of Recognition', *American Journal of Psychology*, 63 (1950), 576-585.
 2. C. H. COOMBS, 'On the Use of Objective Examinations', *Educational and Psychological Measurement*, 13 (1953), 308-310.
 3. B. DE FINETTI, *Op. cit.*, 120.

2. *Autres corrections for guessing*

a. selon l'attrance réelle

Par attrance réelle d'une solution, nous entendons la proportion de sujets qui ont choisi cette solution lors d'une expérimentation précédente¹. L'attrance réelle est souvent bien différente de l'attrance théorique (c'est-à-dire $1/r$). On peut, par exemple, ajouter 1 point par bonne réponse et, pour une réponse fautive, on peut

$$\text{retirer : } 1 \text{ point} \times \frac{(100 - \text{ARBS}) - \text{ARS}}{100 - \text{ARBS}}$$

où ARBS = attrance réelle de la bonne solution, exprimée en centage.

ARS = attrance réelle de la solution envisagée, exprimée en pourcentage.

L'espérance mathématique ne vaut plus 0. On retire 1 point quand l'ARS vaut 0 et on ne retire aucun point quand l'ARS vaut $100 - \text{ARBS}$.

Ce type de formule est indulgent pour les élèves qui choisissent une mauvaise solution fort attractive. Elle paraît plus valable qu'une formule basée sur l'attrance théorique. Cependant, elle suppose que l'échantillon qui a servi à mesurer l'attrance réelle est équivalent à celui auquel on propose la question.

b. selon l'attrance observée sur une population ignorante

L. D'HAINAUT² a posé des questions à choix multiple à des élèves ignorant la matière sur laquelle ces questions portaient. Ses résultats montrent que les choix ne se font pas au hasard, donc que l'attrance réelle est une notion complexe constituée, entre autres, de l'attrance observée sur une population ignorante.

Paradoxalement, on pourrait se demander si toutes les populations ignorantes se valent. On peut éviter ce problème en utilisant comme

1. Voir H. CHERNOFF, 'The Scoring of Multiple Choice Questionnaires', *Amer. Math. Statist.*, 33 (1962), 375-393.

2. L. D'HAINAUT, *L'enseignement des concepts scientifiques et techniques à l'aide de cours programmés* (thèse de doctorat en Sciences Pédagogiques, Université libre de Bruxelles, 1971).

population ignorante les futurs sujets de l'expérimentation. L. D'HAINAUT a développé une formule de cotation basée sur cet indice¹.

B. PRATIQUES INSPIRÉES DU MODÈLE 2

Dans ces pratiques, on demande à l'élève de rejeter certaines questions et d'en choisir d'autres. B. DE FINETTI² cite les procédures suivantes :

1. Barrer autant de solutions que l'on désire, avec un maximum de $r-1$ solutions barrées : c'est la solution proposée par COOMBS, MILHOLLAND et WOMER³ pour évaluer ce que ces auteurs appellent « information partielle ».
2. Choisir une solution (ou aucune) et barrer autant de solutions que l'on désire.
3. Choisir r solutions, en rejeter r' .
4. Choisir et barrer librement les solutions.

Examinons de plus près la procédure de COOMBS, MILHOLLAND et WOMER : ils accordent 1 point par solution fausse éliminée et $(1-r)$ points si la solution correcte est éliminée. Le score à une question peut donc prendre des valeurs allant de $-(r-1)$ à $+(r-1)$, en passant par 0. Un élève qui barrerait toutes les solutions proposées obtiendrait précisément 0, tout comme l'élève qui omettrait simplement la question.

« Contrairement à COOMBS, nous ne considérons pas cela équivalent au rejet de toutes les solutions »⁴. Nous sommes de l'avis de DE FINETTI : l'élève doit pouvoir rejeter toutes les solutions proposées. et certaines questions à choix multiple ne doivent proposer que des solutions fausses. La procédure 4 nous paraît de loin préférable à toutes les procédures vues jusqu'ici.

1. Sera publié sous peu.

2. DE FINETTI, *Op. cit.*, 98.

3. C. H. COOMBS, J. E. MILHOLLAND, and F. B. WOMER, 'The Assessment of Partial Knowledge', *Educational and Psychological Measurement*, 16 (1956), 13-37.

4. DE FINETTI, *Op. cit.*, 96.

C. PRATIQUES INSPIRÉES DU MODÈLE 3

« C'est seulement la probabilité subjective qui peut donner une signification objective à chaque méthode de réponse et de cotation... Des consignes telles que 'choisissez une solution seulement si vous savez qu'elle est juste' ou 'barrez-la seulement si vous savez qu'elle est fausse' sont inévitablement ambiguës précisément à cause de leur précision apparente, qui est si absolue qu'elle est illusoire... »

L'information partielle existe, la détecter est intéressant, nécessaire et possible. L'utilisation de telles méthodes, y compris la façon de coter, et non seulement le système des réponses, doivent être choisis de façon appropriée par l'expérimentateur et expliqués clairement aux sujets qui doivent comprendre la nature du jeu auquel ils jouent. Si cela est fait, les questions au sujet du 'guessing' disparaissent »¹.

DE FINETTI énumère les procédures suivantes :

1. Choisir avec possibilité de mettre un « accent » supplémentaire dans son choix.
2. Accorder à chaque solution une probabilité (multiple de $1/n$) d'être correcte, avec la somme de ces probabilités valant 1.
3. Le système des cinq étoiles² que l'on peut distribuer sur une ou plusieurs solutions. Il n'y a que 7 façons de répartir ces 5 étoiles : 5, 4-1, 3-2, 3-1-1, 2-2-1, 2-1-1-1, 1-1-1-1-1. S'il n'y a que deux, trois ou quatre solutions proposées, certains des derniers arrangements sont impossibles. On peut aussi n'attribuer aucune étoile pour exprimer l'indifférence totale. Des procédures dérivées permettraient à l'élève de distribuer une à dix étoiles et même aucune.

Comme exemple de la procédure n° 2 ci-dessus, J. D. BAKER³ décrit une expérience réalisée au moyen d'un terminal CRT (*cathodic ray tube*) dans le cadre d'un cours ramifié géré par ordinateur. Dans ce système, l'étudiant peut distribuer la totalité (100 %) de sa réponse sur plusieurs solutions. Par exemple, si on propose quatre solutions, un élève peut répondre 5 %, 50 %, 0 %, 45 %, et le total doit valoir 100 %.

1. DE FINETTI, *Op. cit.*, 111.

2. DE FINETTI, *Op. cit.*, 102.

3. J. D. BAKER, 'The Uncertain Student and the Understanding Computer', in *La recherche en enseignement programmé, tendances actuelles* (Paris, Dunod, 1969), pp. 303-319.

Cette procédure très nuancée permet d'obtenir des renseignements sur l'attribution de chaque solution proposée, à partir de très peu d'élèves.

La technique de BAKER est une application des propositions de B. DE FINETTI, et nous regrettons que, comme celles-ci, elle ne permette pas le rejet au même titre que les choix. En outre, la manipulation d'une échelle de pourcentage donne une précision illusoire. Nous avons observé que, dans cette situation, le sujet simplifie systématiquement l'échelle¹, ce qui confirme les exemples donnés par BAKER lui-même². En outre, si après avoir répondu la somme des pourcentages ne vaut pas 100 %, l'élève se débarasse sur n'importe quelle solution des pourcentages restant, ce qui est un artefact regrettable. Enfin, le sujet déplore de ne pouvoir rejeter des solutions, la cote minimale (0 %) représentant l'indifférence.

L'indice de certitude dans la réponse

Nous avons imaginé une procédure où l'élève peut choisir (+), rejeter (−) ou omettre (0) chacune des solutions proposées et où il doit indiquer sa certitude dans son choix ou son rejet par 1 (certitude faible), 2 (certitude normale) ou 3 (certitude forte).

Par exemple, les réponses de trois élèves à une question (où $r = 4$) pourraient être :

N° de la solution	Réponses			Clé de correction ³
	Elève 1	Elève 2	Elève 3	
1	− 3	0	+ 3	+ 1
2	− 3	0	− 3	− 1
3	− 2	+ 1	− 2	+ 1
4	− 3	− 1	− 3	− 1

1. Il utilise très fréquemment des réponses du type 25 %-75 %, ou 90 %-10 %, ou 99 %-1 %, ou encore 50 %-50 %.

2. Les seuls exemples que BAKER fournisse dans son article sont deux démonstrations de réduction d'échelle.

3. Les solutions sont codées par + quand elles sont bonnes et par − quand elles sont fausses. Dans notre exemple, il y a deux solutions correctes, mais il y aurait pu n'y en avoir aucune, ou au contraire il aurait pu y en avoir quatre.

L'élève 1 n'est satisfait d'aucune des quatre solutions proposées : il les rejette avec certitude. L'élève 2 omet de répondre et, quand il répond, c'est de façon très timide. L'élève 3 opère des choix et des rejets très tranchés.

Cette procédure permet la distribution des réponses de l'élève sur toutes les solutions proposées ; elle fait la distinction entre le refus et l'omission et, enfin, elle permet d'obtenir un indice de la certitude de l'élève dans sa réponse. Les chiffres 1, 2 et 3 sont utilisés dans la même acception pour toutes les questions à choix multiple, quel que soit le nombre de solutions proposées, et pour les questions ouvertes. L'emploi de ce système est pratique à la fois pour l'élève qui répond et pour l'expérimentateur qui perfore ces réponses sur cartes¹.

La correction et la cotation des élèves

Pour chaque solution proposée, les cotes peuvent aller de −3 à +3 si nous les calculons par la formule : (réponse × clé). Pour l'ensemble de la question, il est donc possible d'obtenir des cotes allant de −3r à +3r². Cet éventail de cotes résout à sa façon le problème du *guessing*.

Nous présentons aux élèves le tableau suivant :

Bonne réponse et certitude 3		+ 3 points
2		+ 2 points
1		+ 1 point
Omission		0
Mauvaise réponse et certitude 1		− 1 point
2		− 2 points
3		− 3 points

Nous expliquons aux élèves qu'il est préférable d'obtenir 0 plutôt que −2 ou −3, mais que le système n'interdit aucun comportement

1. Afin de n'utiliser qu'une colonne par réponse, on perfore $x + 4$ (où x = la réponse de l'élève), soit un digit de 1 à 7.

2. Avec $r = 5$, il y a 31 cotes possibles ; avec $r = 4$, 25 cotes possibles ; avec $r = 3$, 19 cotes possibles ; avec $r = 2$, 13 cotes possibles.

entraînant une cote défavorable. Très souvent, ce n'est qu'après avoir vu les résultats d'un premier essai que les élèves comprennent que la franchise est la solution la plus payante.

Les réponses sont nuancées, car une certitude 1 n'est pas très éloignée d'une omission. Par contre, un élève tout à fait certain de sa réponse indiquera la certitude 3 ; il risque de gagner ou de perdre 3 points. Nous sommes tout à l'opposé de l'omission : comme dans la vie, la notion de responsabilité est associée à celle de risque. Les *corrections for guessing* classiques pénalisent *également* tous les sujets, alors que certains devinent moins que d'autres. Il y a donc inévitablement hyper-correction ou hypocorrection. Notre système permet d'éviter cet écueil.

L'évaluation des solutions proposées et des questions

L'indice de certitude dans la réponse et la procédure de réponse permettent une analyse plus nuancée de chaque question.

A. Pour chaque solution de chaque question, on peut calculer :

1. Un indice de choix, par exemple le pourcentage des sujets qui ont choisi cette solution et, symétriquement, un indice de rejet, complétés d'un indice d'omission.
2. Des indices de certitude, par exemple le pourcentage (calculé sur le maximum 3) des certitudes exprimées dans le choix et le rejet (qui peuvent aussi faire l'objet d'une ventilation allant de -3 à +3).
3. Des indices (globaux) d'accord avec la bonne réponse, par la formule

$$\text{I.A.S.} = \frac{\sum_i C_i}{3n} \quad \text{où} \quad \begin{array}{l} C = \text{cote d'un élève à la} \\ \text{solution} \\ n = \text{nombre d'élèves de} \\ \text{l'échantillon} \\ \text{I.A.S.} = \text{indice d'accord d'une} \\ \text{solution} \end{array}$$

Ce dernier indice varierait de -1 à +1 (ou de -100 à +100 si on préfère l'exprimer en pourcentage).

4. Une corrélation entre la cote obtenue à la solution et la cote obtenue à l'ensemble de la question, ou à l'ensemble des autres solutions de la question.

B. Au niveau de la question, on peut calculer un indice d'accord par la formule

$$\text{I.A.Q.} = \frac{\sum_j \sum_i C_{ij}}{3(n \cdot r)} \quad \text{où} \quad \begin{array}{l} \text{I.A.Q.} = \text{indice d'accord d'une} \\ \text{question} \\ C = \text{cote d'un élève à une} \\ \text{solution} \end{array}$$

La corrélation entre la cote obtenue à la question et la cote obtenue à l'ensemble de l'épreuve ou à l'ensemble des autres questions de l'épreuve.

Pour construire une question à choix multiple où $r = 4$, on pourrait présenter la question avec $r = 10$ à un petit nombre d'élèves¹, ne retenir que les trois distracteurs les plus pertinents. La simplicité des calculs met cette technique à la portée de tout enseignant.

Dans une optique non plus de construction de questions, mais d'exploitation de celles-ci dans un système de banque², les procédures de réponse et de correction deviennent fastidieuses et, dans une large mesure, inutiles.

Aussi avons-nous développé une procédure abrégée beaucoup plus maniable qui a déjà été appliquée effectivement durant plusieurs mois³. Les résultats obtenus sont encourageants. Les renseignements recueillis sur les élèves et sur les questions sont désormais plus précis et plus fiables ; de plus, la procédure a été rapidement comprise et acceptée par les élèves et les professeurs.

Cette procédure abrégée, son application, les résultats obtenus et le programme d'ordinateur conçu pour la mettre en œuvre feront l'objet d'une autre publication⁴.

Nous voudrions conclure en signalant que cette procédure abrégée cadre parfaitement avec les souhaits de DE FINETTI⁵ :

1. Trente élèves suffiraient.
2. D. LECLERCQ, « Une banque de questions pour l'enseignement », *Éducation* (novembre-décembre, 1971), 67.
3. A l'École Technique de la Force Aérienne Belge à Saffraenberg. L'ensemble des dossiers des élèves et des questions sont traités par un programme d'ordinateur écrit en FORTRAN IV pour des équipements IBM de la série 360. En trois mois, le système a traité 50.000 réponses et leur degré de certitude.
4. Titre provisoire : « L'introduction d'un indice de certitude dans l'évaluation des connaissances ».
5. DE FINETTI, *Op. cit.*, 89.

« La philosophie sous-jacente à toute l'approche est

1. La méthode de cotation aussi bien que les modalités de réponse permises doivent être connues des sujets... qui doivent comprendre pleinement leurs implications notamment en cas d'incertitude dans la réponse.
2. Le sujet doit avoir été intéressé dans l'obtention d'un score total élevé et dans la maximisation de ce score.
3. Les élèves doivent être entraînés à traduire leurs certitudes en probabilités numériques ».

RÉSUMÉ

B. CHOPPIN a décrit trois modèles de l'activité de l'élève ignorant la réponse à une question à choix multiple. Modèle 1 : il choisit une réponse au hasard. Modèle 2 : il élimine les solutions qu'il sait être fausses, puis choisit au hasard parmi celles qui restent. Modèle 3 : il attribue une probabilité d'exactitude à chaque solution, puis il choisit celle dont la probabilité est la plus élevée. L'auteur critique diverses procédures d'utilisation des questions à choix multiple inspirées de ces trois modèles.

Il reprend ensuite l'idée de DE FINETTI sur l'évaluation des connaissances partielles, sous la forme d'un indice de certitude associé au choix ou au rejet. Il propose des formules d'application et de correction aussi bien pour mettre une question à l'épreuve que pour son utilisation courante à des fins d'évaluation.

SUMMARY

B. CHOPPIN has proposed three models of the mental activity of the student who ignores the correct response to a multiple choice question. In model 1, the student chooses randomly among the alternatives. In model 2, he crosses out those alternatives he knows to be wrong; then he chooses randomly among the remaining alternatives. In model 3, he attributes to each alternative a probability of correctness, then he chooses the alternative which has the highest probability. The author criticizes various uses of multiple choice questions derived from those three models. Starting from DE FINETTI's idea about partial knowledge evaluation, the author asks the students to allot marks of confidence 1, 2, 3 to their responses. He proposes application rules and scoring formulas for two different purposes : first, to construct and to improve the questions ; second, to use the questions in everyday school practice.

D. A. LECLERCQ est né à Herstal (Belgique), en 1944.

Instituteur et licencié en sciences de l'éducation, il s'est intéressé à l'enseignement programmé, a rédigé divers cours programmés et mis au point une technique originale intitulée PACELBRO.

Dans ce domaine, il a dirigé des séminaires pour enseignants belges et en coopération avec les pays d'Afrique. Consultant UNESCO en Côte d'Ivoire, en avril-mai 1972, il y a dirigé une enquête nationale sur la première année d'Education télévisuelle.

Dans le domaine des questions à choix multiple, il a participé activement, durant deux ans, à la création de banques de questions pour lesquelles il a déterminé les modalités pédagogiques et écrit les programmes FORTAN.