

On se méfie des QCM dans les milieux éducatifs. Et on a raison ! Les dangers sont évidents : tentation pour l'enseignant de se cantonner aux questions de détails (donc, pour l'apprenant de ne plus étudier que ces derniers), déshabitude de l'étudiant à formuler sa pensée, à rédiger ses réponses, renforcement d'une tendance à considérer la connaissance comme tout ou rien.

Or, on peut échapper à ces dangers, à condition de concevoir les QCM selon les principes originaux, comme la sollicitation implicite où l'élève doit découvrir la question cachée dans la question.

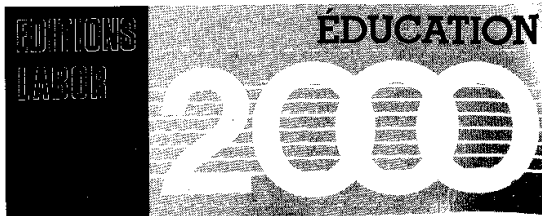
Cet ouvrage propose une classification des QCM basée sur le bon sens, la rigueur... et la recherche en éducation. Il en va de même pour les (20) règles de rédaction, qui sont justifiées par les résultats d'expériences américaines originales.

Enfin, les avantages et inconvénients des QCM sont intégrés dans un modèle de l'évaluation pédagogique.

IMPRIME EN BELGIQUE

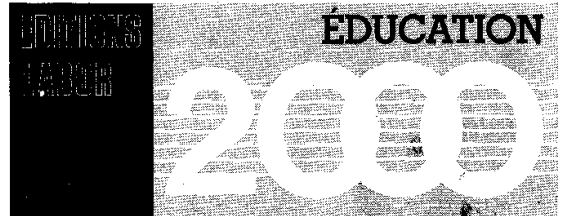
D/1986/258/20

ISBN 2-8040-0154-7



Dieudonné LECLERCO

LA CONCEPTION DES QUESTIONS A CHOIX MULTIPLE



D. LECLERCQ

La conception des questions à choix multiple

Collection
EDUCATION 2000
dirigée par Gilbert DE LANDSHEERE

Editions LABOR, Bruxelles

Avant-propos

Le présent livre est entièrement consacré aux questions à choix multiple c'est-à-dire à une méthode particulière d'évaluation. Il s'adresse à des lecteurs déjà informés des problèmes généraux relatifs aux objectifs pédagogiques et à la docimologie. Il s'inscrit dans le prolongement d'ouvrages de base tels que ceux de V. et G. De Landsheere, *Définir les objectifs de l'éducation* (1982) et de G. De Landsheere, *Evaluation continue et examens, Précis de docimologie* (1980).

Seul le premier chapitre revêt un caractère général : il traite du processus de régulation et de la place qu'y occupe l'évaluation. Le reste des considérations est délibérément technique.

Pourtant notre propos n'est pas destiné aux seuls spécialistes, loin s'en faut. Nous avons voulu donner une information précise et nuancée sur l'un des principaux outils de l'évaluation, en indiquer les grandes possibilités, en expliquer l'utilisation correcte et en montrer les avantages et les limites. C'est donc à tous ceux qui doivent évaluer dans le domaine des sciences humaines que nous avons pensé, plus spécialement aux éducateurs de tous les niveaux et aux étudiants en psychologie et en sciences de l'éducation.

© 1986, Editions LABOR
Chaussée de Haecht, 156-158 - 1030 Bruxelles
D/1986/258/20
ISBN 2-8040-0154-7
L. 90 80 76

Remerciements

Pendant des années, le professeur G. de Landsheere nous a encouragé à poursuivre la rédaction du présent ouvrage et nous a prodigué ses conseils.

Nos collègues, Claude Lutgens, Dominique Lafontaine et Armand Hoyoux nous ont aidé par leurs critiques et leurs suggestions de fond et de forme.

D'autres collègues, inspecteurs, enseignants et chercheurs, nous ont permis d'affiner certaines idées.

Nelly Saenen, avec l'aide d'Anna Charlier, Anne-Marie Di Palma, Carla Englebert et Germaine Nibus, a dactylographié les diverses versions du présent texte.

Notre travail a été grandement facilité par le soutien quotidien des membres de notre famille : Monsieur et Madame Boxus, Monsieur et Madame Duchateau, Monsieur et Madame Leclercq, et surtout Elise Boxus, notre épouse.

CHAPITRE I

Evaluation et questions à choix multiple

A. Définitions et terminologie

1. Les trois composantes des QCM.
2. L'expression QCM.
3. Définition d'une QCM.

B. Modalités de questionnement

1. Modalités d'obtention d'une réponse.
2. Techniques de sollicitation directe.
3. Objectivité de la correction et docimologie négative.
4. Créativité dans l'évaluation et docimologie positive.
5. QCM et histoire des tests.

C. Avantages et désavantages des QCM

1. L'avantage « quantitatif ».
2. Les avantages qualitatifs.
3. La rançon des avantages.
4. Trois désavantages fondamentaux.

D. Le problème de la mémorisation des solutions fausses

1. La critique de Skinner.
2. Les travaux de Preston.
3. Les travaux de Karraker.
4. Discussion.

E. Le problème de l'uniformisation des individus

1. Les travaux de Kendler.
2. Où sont les vrais problèmes ?

F. Synthèse : Régulations, évaluations et QCM

1. Régulation et évaluation pédagogiques.
2. Les avantages et désavantages des QCM dans la régulation.
3. Conclusions.

A. Définitions et terminologie

I. Les trois composantes d'une QCM

Trop souvent, on ne s'attache qu'aux deux parties les plus visibles d'une QCM : l'amorce (ou introduction) et les solutions proposées (ou suggestions). On néglige souvent la consigne. On commet ainsi une erreur aussi grave que celle qui consisterait à réduire un arbre à son tronc et aux branches, en oubliant les racines, les fleurs, les feuilles, les fruits, etc.

La consigne est donnée à l'élève au début de l'épreuve. Elle devrait toujours être écrite. La plupart du temps, elle est valable pour toutes les questions d'une épreuve.

LES TROIS COMPOSANTES

1. Les consignes décrivent :

- la question
- le mode de réponse
- les principes de notation

d) Le barème des conséquences (aussi appelées notes, tarifs, points à gagner et à perdre) doit préciser les notes obtenues en cas de réponse correcte, incorrecte ou en cas d'omission.

2. L'amorce

(aussi appelée tronc, corps, base, article de base, prémisse)

- définit le problème (on veut...).
- pose la question (laquelle...).

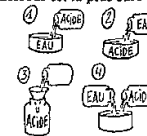
3. Les solutions proposées

(aussi appelées choix, suggestions, options) comprennent la solution correcte (ici 1) et des solutions incorrectes, appelées *distracteurs* (ici au nombre de 3).

EXEMPLE

- Pour les questions qui suivent, une seule solution est considérée comme correcte.
- Choisissez une seule des solutions proposées ou abstenez-vous de répondre (omission).
- Si vous trouvez la solution correcte, vous gagnez des points ; une solution incorrecte vous en fera perdre. Vous avez intérêt à ne rien écrire plutôt que de choisir au hasard.
- La notation se fera selon le barème suivant :
 - Réponse correcte (TC) : + 2 points
 - Réponse incorrecte (TI) : - 1 point
 - Omission (TO) : 0 point

On veut diluer de l'acide sulfurique concentré. Laquelle des manipulations ci-dessous est la plus sûre ?



2. L'expression « Question à choix multiple »

a) Question

Certains auteurs français utilisent le sigle QCM pour signifier *Questionnaire* à choix multiple. Cette dernière appellation ne correspond pas à l'expression anglaise originale (*multiple choice question*) et diffère des expressions utilisées dans les autres langues (ex : *meerkeuze vraag* en néerlandais).

De plus, l'expression Questionnaire à choix multiple convient seulement au cas du test qui ne comporte que ce type de questions.

b) Choix

En français, le mot choix a plusieurs sens. Le dictionnaire *Petit Robert* (1976) en distingue quatre. Deux d'entre eux correspondent au sens du terme choix dans l'expression QCM.

Deux définitions de « choix »	Voir
● Existence de plusieurs partis entre lesquels on peut choisir (passif). Ex. « avoir le choix », « on vous laisse le choix ».	Option Alternative Dilemme
● Ensemble de choses parmi lesquelles on peut choisir. il y a un grand choix de poissons au H.C.B.	Assortiment. Collection. Eventail. Réunion.

c) Multiple

On est souvent tenté de croire que le terme *multiple* évoque, comme le terme *multiplicité*, un nombre considérable, alors qu'il a pour antonymes (selon le *Petit Robert*) les termes *simple*, *un*, *unique*. Le choix « multiple » s'oppose à *alternative* (« choix entre deux possibilités ») ou *dichotomie*.

*
* *

Par commodité, nous avons opté pour l'écriture du sigle QCM sans points.

3. Définition d'une QCM

Toute situation de choix ou toute question fermée peut-elle être appelée « QCM » ? Si l'on s'en tient rigoureusement aux quatre composantes que l'on vient de distinguer, notamment aux barèmes, une *définition* plus restrictive s'impose.

Nous proposons d'appeler QCM...

« Une question à laquelle l'étudiant répond en opérant une sélection (au moins) parmi plusieurs solutions proposées, chacune étant jugée (par le constructeur de l'épreuve et par un consensus entre spécialistes) correcte ou incorrecte indépendamment de l'étudiant qui doit y répondre ».

Ainsi, les trois questions suivantes ne sont pas des QCM parce que, bien qu'une réponse correcte existe, elle n'est pas (encore) disponible à celui qui pose la question :

- A. A l'issue des plaidoiries des avocats, les membres du jury se demandent : « Cet accusé est-il coupable ? » Oui - Non
- B. A l'examen des cours de la bourse, un détenteur de fonds se demande : « De quelle entreprise est-il le plus intéressant d'acheter des actions ? » X ? Y ? Z ?
- C. Un amoureux se demande : « Que répondra Alice à ma demande en mariage ? » Oui ? Non ? Autre chose ?
- D. De même, la question qui suit n'est pas une QCM au *sens strict* parce que la solution correcte varie selon les individus qui répondent.

Etat civil ? Célibataire, marié, veuf, divorcé

Les quatre questions ci-dessous ne sont pas des QCM au *sens strict* parce que la réponse à ces questions est affaire soit d'opinion personnelle, soit de goût. Il ne peut être question d'y associer un barème de notation.

Choix professionnel :

« Quelles études dois-je entreprendre ? » A ? B ? C ?

Choix sentimental :

« Quel garçon dois-je épouser ? » A ? B ? C ?

Choix politique :

« Pour quel parti dois-je voter ? » A ? B ? C ?

Choix artistique :

« Quelle est l'œuvre de Tchaikovsky la plus mélodieuse ? » A ? B ? C ?

Ne sont donc pas des QCM :

- les questionnaires d'opinion où les sujets doivent noter leur avis sur une échelle allant de « Pas du tout d'accord » à « Tout à fait d'accord » ;
- les échelles d'évaluation courantes utilisées pour apprécier des travaux au moyen de notes ordonnées allant, par exemple, de « Insuffisant » à « Excellent ».

B. Modalités de questionnement

I. Modalités d'obtention d'une réponse

La QCM apparaît comme un mode de questionnement très direct, voire brutal. Cependant, on verra (au chapitre 3) qu'elle peut être mise au service de sollicitations plus indirectes de la réponse.

Le comportement de l'étudiant qui fait l'objet de l'évaluation peut avoir été produit dans des circonstances très différentes, selon que :

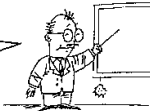
- a) Le professeur a sollicité explicitement la réponse ;
- b) Le professeur a sollicité implicitement un comportement ;
- c) L'étudiant a émis spontanément un comportement.

Ces trois situations sont illustrées ci-après à propos d'un même contenu : la géographie.

Situation a

Figure 1.2

Sachant que les villes de Zürich et Budapest sont situées à la même latitude (47°) et, respectivement, à 9° et 19° de longitude EST, quelle est la distance (en kilomètres) entre ces deux villes ?



Il s'agit là d'une sollicitation explicite : l'élève doit répondre à une question qui lui est posée. C'est le mode de questionnement le plus fréquent dans les écoles.

Situation b

Figure 1.3



On sait que le tour de la terre est de 40 000 km et qu'un tour de la terre correspond à 360° de longitude. Dès lors, le nombre de kilomètres (Nkm) séparant Budapest de Zürich se calcule comme suit :

$$Nkm = \frac{40\,000 \times 10}{360}$$

Le professeur a sciemment introduit une erreur dans son raisonnement. C'est une sollicitation implicite de réponse. Le professeur n'a pas posé la question et ce afin de permettre à l'étudiant de se la poser lui-même et de raisonner, par exemple, de la façon suivante :

Figure 1.4

Le tour de la terre ne vaut 40 000 km qu'à l'équateur. A la latitude de Zürich (47°), un tour de terre est bien plus petit !



Situation c

Figure 1.5

Si on travaillait en grades (qui vont de 0 à 100) plutôt qu'en degrés (qui vont de 0 à 90), les calculs seraient simplifiés.



L'étudiant agit ici spontanément.

Les comportements du type c revêtent une qualité éducative particulière : pour se manifester, l'individu n'attend pas que la question lui soit posée : il n'est même pas nécessaire qu'une situation ait été volontairement aménagée par l'éducateur (comme dans la situation b).

Les comportements spontanés constituent le but ultime de l'éducation ; les autres sont plutôt des moyens d'atteindre ce but. En créant des situations du type b (questionnement implicite), l'enseignant peut favoriser l'apparition de la situation c (réponses spontanées).

Les QCM ne s'appliquent-elles qu'à la situation a ? On verra (au chapitre 4) qu'il existe, grâce aux QCM, une façon économique et systématique (les solutions générales) de créer des situations du type b (sollicitations implicites).

2. Les techniques de sollicitation directe de la réponse

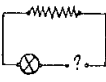
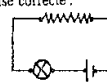
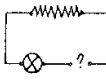




a) Les questions ouvertes et les questions fermées

On peut répartir les questions en deux grands groupes :

- 1° Les questions à *réponses construites* (ou produites, ou rédigées). On parle aussi de *questions ouvertes* pour mettre en évidence la plus ou moins grande liberté (de forme et de fond) dont jouit l'étudiant.
- 2° Les questions à *réponses choisies* (ou sélectionnées). On parle aussi de *questions fermées* pour souligner que le nombre de possibilités de réponse qui s'offrent à l'étudiant est limité.

Voici, sur un même contenu, une question de chaque type et, à droite, les réponses correctes :

Figure 1.6

<p>Q1 Complétez le schéma du circuit électrique suivant :</p> 	<p>Réponse correcte :</p> 
<p>Q2 Voici un schéma électrique incomplet :</p>  <p>Quel élément doit-on introduire pour en faire un circuit électrique complet ?</p> <p>1  2  3  4 </p>	<p>Réponse correcte : 2</p>

Cet exemple illustre qu'en général, pour un même contenu, la question fermée est plus longue que la question ouverte, mais que la réponse à fournir est plus courte. De même, le temps consacré par l'enseignant à la rédaction de la question fermée est plus long tandis que le temps nécessaire à la correction est plus court.

b) Les questions à encadrement ou à intervalle de réponse

Ces questions peuvent exiger de l'étudiant qu'il propose un intervalle (voir exemple A ci-après) ou qu'il choisisse parmi plusieurs intervalles proposés (voir exemple B ci-après).

Exemple A

En quelle année est mort Jules César ?

Répondez par deux dates constituant un intervalle dans lequel se situe à coup sûr la date exacte.

RC = en 44 av. J.-C.

Le score de l'étudiant sera fonction du fait que l'intervalle contiendra (ou non) la date exacte et de l'ampleur de cet intervalle (plus il est petit, plus on gagne de points en cas de réponse correcte).

Exemple B

En quelle année est mort Jules César ?

1. Entre 50 et 40 av. J.-C.
2. Entre 60 et 30 av. J.-C.
3. Entre 40 et 20 av. J.-C.
4. Entre 60 et 44 av. J.-C.

Les deux premières solutions sont correctes et les deux dernières incorrectes. La première, plus « précise » (intervalle plus étroit) que la deuxième doit rapporter plus de points car le risque pris est plus grand¹.

c) Une classification des principaux types de questions

Parmi les questions ouvertes, on peut distinguer trois sous-catégories :

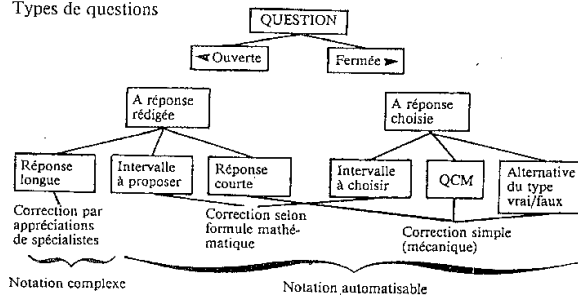
- Les questions à réponses longues et libres (rédaction, rapport d'expérience, projet de recherche, exposé oral, etc.) ;
- Les questions où l'étudiant doit proposer un intervalle ;
- Les questions à réponses courtes et de forme imposée (un mot, une seule phrase, un nombre, une date, un verbe, un nom de personne, une formule mathématique, un alexandrin, etc.).

¹ La formule de calcul qui permet de tenir compte de ces éléments est assez complexe. On peut, par ailleurs, regretter que ce type de question (invitant à proposer des intervalles) soit beaucoup trop peu utilisé dans les écoles. Hardy (1979 et 1981) a réalisé une importante recherche sur ce thème et propose des formules de score et des programmes d'entraînement par ordinateur pour promouvoir ce type de question.

Parmi les questions fermées, on distingue trois formes :

- Les questions à intervalles proposés, QCM dont les tarifs sont calculés, solution par solution, au moyen d'une formule spéciale.
- Les QCM classiques,
- Les alternatives, qui présentent deux solutions à l'exclusion de toute autre, et sont de la forme VRAI-FAUX (OUI-NON, EXACT-INEXACT, EST-N'EST PAS,...). Ce sont des QCM si particulières qu'elles constituent un groupe à part des QCM classiques.

Figure 1.7
Types de questions



Les réponses aux questions fermées sont *précodées* : l'étudiant choisit des numéros ou des lettres correspondant aux solutions qu'il retient. La réponse courte aux questions ouvertes peut être *post-codée* par le professeur (par exemple, 1 en cas de réussite, 2 en cas d'échec, 0 en cas d'omission). Les réponses consistant à proposer un intervalle ne peuvent être codées, mais bien standardisées.

Il est dès lors possible de faire traiter toutes ces réponses par un programme d'ordinateur, ce qui présente bien des avantages. Un traitement informatique permet, en effet, non seulement de réduire le coût et le délai de la correction, mais aussi de développer des procédures de diagnostic, de remédiation, de formation, qui auraient été impossibles autrement.

Les questions ouvertes à réponse longue (rédaction, dissertation, rapport d'expérience, projet de recherche,...) entraînent généralement des corrections complexes, où des éléments de réponses sont imbriqués les uns dans les autres. Dans ce cas, le correcteur doit faire un important effort d'analyse

pour séparer a posteriori les éléments de la réponse relevant de catégories différentes. Par exemple, au lieu de poser 20 questions fermées et d'obtenir 20 réponses, donc 20 notes, on pose une question, on obtient une réponse longue et on la note à 20 points de vue différents. De Bai, De Landsheere et Beckers (1977) ont proposé des échelles d'évaluation descriptives adaptées à ce type de correction.

3. Objectivité de la correction et docimologie négative

Une part du succès rencontré par les QCM vient de ce qu'elles échappent à la subjectivité de la correction que l'on déplore dans d'autres modes d'interrogation.

Pour garantir l'objectivité de la mesure dans une épreuve pédagogique, on a d'abord songé à la *standardiser*², c'est-à-dire à rendre son contenu et ses modalités d'administration identiques pour tous les étudiants. C'est le principe des examens externes³, où les questions sont conçues pour des ensembles locaux, régionaux⁴, ou nationaux⁵.

Cette précaution est nécessaire pour garantir l'objectivité de la mesure, mais elle est insuffisante. Pieron et ses collaborateurs l'ont montré dans leurs études sur le baccalauréat français. Ils ont essentiellement dénoncé l'absence de *standardisation dans les critères de notation* et, par conséquent, l'absence de fidélité de la mesure. Ces travaux ont constitué, en France, le point de départ de la science des examens, la docimologie (De Landsheere, 1977).

Pieron (1969, p. 105) cite les indices de corrélations moyens entre six correcteurs évaluant les mêmes travaux :

- Dissertation philosophique	0,447
- Composition française	0,549
- Anglais	0,797
- Version latine	0,833
- Physique	0,863
- Mathématique	0,888

Plus loin, Pieron (1969) commente ainsi l'une de ses observations :

« Dans les couplages de correcteurs au diplôme d'Etat de conseiller d'orientation, j'ai obtenu 27 coefficients de corrélation, dont la moyenne générale était de + 0,506, mais compris dans une marge allant de 0,070 à 0,890. C'est dire qu'un couple d'examineurs peut quelquefois arriver à

² Voir De Landsheere (1979).

³ Par opposition aux examens internes, conçus par les enseignants de l'école que fréquente l'étudiant.

⁴ Les examens cantonaux en Belgique, par exemple.

⁵ Le baccalauréat en France, par exemple.

un certain accord de classement, ou peut au contraire se comporter comme si chacun fixait sa note à peu près au hasard » (p. 144).

Quand plusieurs correcteurs aboutissent aux mêmes notes pour une même copie, on parle de *concordance* (intercorrecteurs).

Quand un correcteur aboutit aux mêmes notes, lors de deux corrections différentes des mêmes copies, on parle de *constance* (intracorrecteur). De Landsheere (1976, p. 36) signale l'expérience suivante : « Quatorze historiens ont été invités à noter une deuxième fois quinze compositions, douze à dix-neuf mois après les avoir cotées une première fois. Toute trace de correction avait été effacée. Les professeurs accordaient non seulement des points, mais indiquaient la réussite globale ou l'échec. Dans nonante-deux cas sur deux cent dix, le verdict a été différent d'une fois à l'autre. Il faut toutefois insister sur le fait que des résultats aussi pauvres sont dus au manque de directives rigoureuses précisant les aspects à considérer par les notateurs ».

Pieron appelle « attitude typologique d'un correcteur » un ensemble de caractéristiques présentées par les notes qu'il accorde.

- Le centrage ou la moyenne de ses notes (certains examinateurs se révèlent trop sévères et d'autres excessivement généreux) ;
- L'ampleur de la distribution de leurs notes (dispersion exprimée par l'écart-type) ;
- La forme de la distribution de leurs notes (normale, symétrique ou non) ;
- La constance, dans le temps, de ces trois paramètres.

Les six effets de correction les plus connus sont les suivants :

a) *Effets liés principalement au seul correcteur.*

1° *Effet de sévérité* : certains correcteurs sont systématiquement plus sévères que d'autres. Pieron (1969, p. 24) remarque que dans des cas extrêmes, « pour prédire la note d'un candidat, il vaut mieux connaître son examinateur que lui-même ».

2° *Effet de tendance centrale* : la majorité des correcteurs évitent les extrêmes, concentrent leurs notes au milieu de l'échelle et donnent à leur distribution une forme gaussienne.

b) *Effets liés à la relation entre le professeur et un élève particulier.*

3° *Effet de halo* : la note est influencée (« contaminée ») par des caractéristiques de l'étudiant, étrangères à sa performance : aspect physique, présentation vestimentaire, prononciation, accent, etc. Ebel (1965, p. 183) note que des contre-performances lors de tests peuvent être révélatrices d'évaluations surfaites lors de situations non standardisées.

4° *Effet de stéréotype ou d'inertie* : le correcteur a tendance à attribuer à un étudiant les notes que celui-ci a acquises auparavant.

c) *Effets liés à la correction de plusieurs copies :*

5° *Effets de séquence et de contraste* : la copie qui suit une copie brillante risque d'être désavantagée et inversement.

6° *Effet de relativisation* : chaque copie est jugée par rapport aux autres copies. Parmi des copies qu'il a jugées toutes « moyennes » quelques mois auparavant, un professeur distinguera, lors d'une seconde correction, des « faibles » et des « bonnes ».

Augmenter le nombre de correcteurs permet, en général, de combattre ces « effets ».

« Etant donné le fait de la fluctuation des notations relevant de ce facteur aléatoire qu'est la personnalité du correcteur, on est en droit de penser qu'en multipliant les correcteurs, on compensera ces fluctuations. Laugier et D. Weinberg ont appelé valeur « vraie » la moyenne d'un nombre assez grand de notations indépendantes. Ils ont cherché à déterminer le nombre minimum d'examineurs compétents auxquels il faudrait faire appel pour obtenir la notation méritant confiance. Dans ce but, ils ont utilisé une formule statistique relative à l'augmentation de la fidélité des mesures en fonction de leur nombre et de leur degré de concordance.

Recherchant un coefficient élevé de fidélité (0,99), et se fondant sur les moyennes des indices de corrélation obtenus pour chaque catégorie d'épreuves, ils ont trouvé que ce nombre minimum était le suivant :

Composition française	78
Version latine	19
Anglais	28
Mathématique	13
Dissertation philosophique	127
Physique	16

(Pieron, 1969, p. 23).

Mais pareille solution ne manque pas de lourdeur, et l'on comprend que l'objectivité de la correction d'une QCM soit un avantage tant apprécié en docimologie.

4. Créativité dans l'évaluation et docimologie positive

Les travaux de Pieron et ses continuateurs sont peut-être l'exemple le plus frappant d'une docimologie négative, qui étudie les dysfonctionnements d'un système sans y porter remède. La docimologie positive consiste

à créer des méthodes, des techniques et des instruments nouveaux (dont les QCM) adaptés aux problèmes. Des utilisations nouvelles des QCM constituent un apport à la docimologie positive.

Mais, ce que l'on veut gagner en objectivité, ne va-t-on pas le perdre en appauvrissant la mesure ? Autrement dit, les QCM sont-elles moins *valides* que les questions ouvertes ? Ces dernières représentent-elles la forme parfaite de questionnement dans tous les domaines et pour tous les comportements et les autres formes (QCM, intervalles) n'en sont-elles qu'un ersatz inévitable, qu'un substitut qu'il faut bien supporter ?

Cette croyance est fort répandue et renforcée par certaines utilisations peu heureuses des QCM. Un examen de situations particulières oblige à plus de nuances. Ainsi, pour mesurer « la connaissance de la chronologie des faits historiques », est-il préférable d'exiger une date précise ? (« En quelle année a eu lieu... » ? exercice de pure mémoire) ou d'inviter les étudiants à répondre par un intervalle ? Cette dernière formule favorise le raisonnement historique, basé sur une exploitation efficace d'un petit nombre de dates connues (repères) et d'un grand nombre de faits en relation entre eux. Cet « argument de pertinence » illustré ici par les réponses par intervalles sera souvent repris ci-après à propos des QCM.

Dans le domaine de l'évaluation, l'heure est moins aux bilans qu'aux innovations. Les QCM se révèlent un instrument d'une richesse insoupçonnée, à condition d'en exploiter l'infinie variété. Nombre de leurs avantages sont déjà connus (voir section D ci-après), mais beaucoup sont encore à découvrir.

5. QCM et histoire des tests

Depuis l'invention des tests, les QCM connaissent, aux Etats-Unis, un succès bien plus grand qu'en Europe. Il y a à cela deux raisons principales.

La première est qu'aux Etats-Unis, les tests jouent un rôle important dans la pratique des examens. Ainsi, les universités acceptent les étudiants sur la foi non seulement des résultats obtenus dans l'enseignement secondaire, mais aussi des scores obtenus à des épreuves standardisées.

Pour se préparer à ces tests, les candidats trouvent en librairie les versions des années précédentes, avec réponses correctes et commentaires. Ce genre de préparation est là-bas tout à fait courant.

La deuxième raison est plus subtile. Comme l'explique Popham (1980), la confiance accordée aux QCM n'est pas sans liens avec la première guerre mondiale et son issue. Les Etats-Unis ont été confrontés aux problèmes de sélection des futurs officiers et sous-officiers parmi les mobilisés. Il fallait

faire vite. Sur la base de travaux de psychométriciens de l'époque (Goddard, Otis, Bingham, Wells, Thurstone, etc.), furent créées les diverses formes des célèbres « Army Tests » (Army Alpha Examinations), composés de QCM. L'opération fut menée à bien et les Américains contribuèrent manifestement à la victoire des alliés. Les QCM avaient ainsi gagné leur titre de noblesse.

C. Avantages et désavantages des QCM

On ne peut faire un bilan objectif des avantages et des désavantages des QCM. Tout dépend de la façon dont on exploite les premiers et dont on minimise ou contourne les seconds, ainsi que du système de valeurs de la personne qui fait ce bilan.

1 Les avantages « quantitatifs » des QCM

Les QCM contribuent en principe à améliorer la validité, la fidélité et la sensibilité d'un examen parce qu'elles constituent un moyen d'obtenir et de corriger un grand nombre de réponses en peu de temps.

Mais cette fonction ne peut être remplie que par de bonnes QCM qui présentent les trois caractéristiques suivantes :

- a) la question est claire (ce qui facilite la compréhension) ;
- b) la réponse est brève (ce qui évite à l'étudiant une formulation laborieuse) ;
- c) la correction est simple (ce qui évite au professeur des prises de décision fastidieuses).

a) La clarté de la question

Les termes utilisés dans l'énoncé d'une QCM doivent être choisis avec soin afin de rendre la question aussi précise que possible. Une même attention sera portée à la syntaxe, à la structure de la question (voir chapitre 3) et au choix des solutions proposées.

L'éducateur qui rédige des QCM trouvera avantage à les soumettre à des collègues avant de les essayer avec des élèves. Les réactions de ces derniers sont toujours riches d'enseignement et font d'habitude l'objet d'analyses statistiques appropriées.

La mise au point expérimentale de QCM nécessite parfois plusieurs applications préliminaires suivies d'analyses statistiques et remaniements successifs. Moyennant des précautions et avec de l'expérience, on arrive à construire des QCM d'une grande limpidité, ce qui permet à l'étudiant de consacrer l'essentiel de son temps à la réponse. Le chapitre 3 présente des règles de création et de rédaction des QCM.

b) La brièveté de la réponse

L'élève ne doit pas rédiger une réponse, mais indiquer ou recopier une lettre ou un chiffre. Pour s'exprimer, l'étudiant est ici dispensé de la recherche d'un vocabulaire ou d'un style adéquat : il est débarrassé du souci de l'orthographe et d'une syntaxe appropriée. Il n'a même plus à se demander si son écriture est lisible ou non.

Or, il n'est pas exceptionnel que la difficulté d'exprimer correctement une réponse (connue par l'élève) affecte la mesure. Vandeveldt (1971) regrette, par exemple, que l'analyse littéraire « se juge exclusivement sur la production et même sur la *formulation de cette production* » qui parasite la performance que l'on veut observer ».

c) La simplicité de la correction

Fréquemment, on demande à l'étudiant d'inscrire ses réponses aux QCM dans une grille imprimée sur une bandelette de papier. Dans l'exemple ci-dessous, le professeur a placé sa clé de correction sous les réponses de l'élève Dupont et a barré d'un trait oblique les réponses incorrectes.

Nom : Dupont

Questions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Réponses de l'élève	A	2	2	2	A	3	1	7	A	3	A	1	2	5	3	1	3	5	6	A	2	A	5	7	5	3				
Clé de correction	3	5	1	2	2	4	3	1	7	3	3	3	2	1	2	4	5	3	1	3	5	6	1	2	2	4	5	6	5	3

Des analyses complémentaires peuvent être aisément réalisées. Par exemple, après la première correction ci-dessus, une seconde grille permet de ne faire apparaître que les questions mesurant la « connaissance de mémoire ».

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	

On constate ici que, pour cette catégorie de questions, l'élève Dupont a fourni huit réponses correctes et quatre incorrectes.

Un tel système permet à un enseignant de fournir, dans l'heure, les scores (détaillés) obtenus par une trentaine d'étudiants à une trentaine de questions.

Pour couvrir une matière en peu de temps, obtenir des sous-scores permettant des diagnostics précis (tests formatifs), calculer des scores totaux valides et fidèles (tests de classement), les QCM sont particulièrement indiqués.

d) La capacité de couvrir une matière

Cet avantage est la conséquence des avantages signalés ci-dessus. On connaît trop l'angoisse des étudiants lorsque leur connaissance d'une vaste matière est évaluée au travers de trois ou quatre questions seulement, tirées au hasard ou, plus grave encore, posées selon l'inspiration du moment. L'interrogation par un grand nombre de QCM garantit mieux la représentativité de l'échantillon des questions et donc une meilleure évaluation des connaissances réelles de l'étudiant. Ce sentiment de justice est dû à la validité, la fidélité et la sensibilité de la mesure que la seule augmentation du nombre de questions ne suffit pas à assurer. La qualité psychométrique des questions doit aussi être établie.

2. Les avantages qualitatifs des QCM

Une QCM présente certaines caractéristiques qui lui donnent des avantages qualitatifs :

- Certains niveaux supérieurs d'activité mentale peuvent être évalués de façon systématique et précise.
- Les consignes mentionnent clairement les exigences du professeur quant à la réponse à fournir.
- La correction est simple, automatique et objective.
- La concision des réponses permet les analyses synoptiques et l'autorégulation.

Examinons ces quatre aspects de façon plus détaillée.

a) La possibilité d'évaluer systématiquement, et précisément, des niveaux supérieurs d'activité mentale

De trop nombreuses QCM existantes portent sur des points de détail (voir section 3c) ou ne mettent en jeu que des processus intellectuels élémentaires, la mémoire, par exemple... et même de préférence la mémoire de reconnaissance plutôt que la mémoire d'évocation.

Mais il n'y a là rien d'inhérent aux QCM et cet ouvrage précisément a pour but de le démontrer. Contrairement à la première impression qu'elles donnent, les QCM peuvent être un précieux adjuvant pour évaluer des niveaux supérieurs d'activités mentales. Pour ce faire, il faut néanmoins réunir certaines conditions. Ainsi, plutôt que de se limiter à la *consigne classique* et à la forme habituelle d'une QCM, on puisera plus largement dans la gamme des possibilités (voir chapitres 2 et 4). Après avoir lu le chapitre 4, essayez donc de répondre aux quinze questions de la section E, afin de vivre « de l'intérieur » (c'est-à-dire du point de vue de l'étudiant) une évaluation systématique de la vigilance cognitive, de la logique, ainsi que de la compréhension profonde de principes et de concepts.

b) La précision des exigences

La question *ouverte* ne pousse pas toujours l'étudiant « dans ses derniers retranchements ». C'est ce qu'expriment Noizet et Caverni (1978) en remarquant qu'une QCM « ne permet pas l'esquive, à la différence des épreuves dites de production qui permettent toujours au candidat de masquer une absence de savoir » (p. 186), car « en général, le candidat parle de ce qu'il sait et évite de parler de ce qu'il connaît moins bien » (p. 172). Cette impossibilité de contourner la difficulté explique probablement certaines réticences des étudiants vis-à-vis des QCM.

Hallux-Hendrix (1969) observe que les questions ouvertes donnent souvent lieu à la répétition de la réponse apprise en classe, *sans variante*. Elle a montré (voir chapitre 4 section B) en quoi les QCM permettent de confronter les élèves à des aspects d'une matière qui n'ont pas forcément été vus en classe, ou de mettre en évidence des lacunes et oublis de l'enseignement (qui peut, dès lors, faire l'objet d'une régulation).

c) L'objectivité de la correction

Comme les critères de correction sont définis à l'avance, les biais que pourraient introduire les correcteurs sont éliminés ; cette caractéristique contribue à rendre représentatives les mesures-échantillons (opération 4 de la régulation ; cf. section F), donc à augmenter la fidélité.

Cet avantage favorise aussi l'autorégulation : les étudiants trouvent ici le moyen de s'évaluer rigoureusement eux-mêmes.

On peut par exemple demander aux étudiants de répondre sur deux bandellettes superposées (celle du dessous est « autocopiante »). Un exemplaire est remis au professeur qui le corrigera ou le fera corriger de façon approfondie, par ordinateur. L'autre exemplaire peut faire l'objet d'une correction immédiate, collective ou individuelle, en classe ou à domicile.

La discussion « à chaud » des erreurs offre des possibilités d'apprentissage non négligeables : on sait que l'information en retour (feedback) est, en général, d'autant plus efficace qu'elle est rapide.

d) La concision du code permettant les analyses synoptiques et l'auto-régulation

Un *code bref* (une seule lettre ou un seul chiffre) permet au professeur de prendre connaissance simultanément et synoptiquement de la réponse de chaque élève à une QCM, pendant le déroulement même de la leçon.

La procédure la plus élémentaire consiste à demander à l'élève de montrer le chiffre-réponse à la manière des juges des épreuves de patinage artistique. Une technique plus complexe utilise un clavier ad hoc : l'ensemble des réponses apparaît au professeur sur un « tableau de bord » et peut faire l'objet de traitements statistiques ou autres. Le professeur, devant les résultats de *tous* les élèves, peut prendre des décisions sur le champ.

Le « *précodage* » facilite également la correction par l'élève. Des dispositifs mécaniques, optiques, chimiques, électriques, audio-visuels et informatiques offrent diverses possibilités d'auto-correction immédiate. L'automatisation possible de l'analyse est parfois considérée comme un avantage mineur des QCM. Mais, avec les développements de l'informatique, elle constitue une source réelle de progrès, tant pour la pratique scolaire que pour la recherche.

*

* *

En résumé, la précision des exigences permet de disposer, pour chaque étudiant, d'informations diagnostiques systématiques, qui rendront la régulation efficace. Outre qu'elle affecte favorablement la fidélité des mesures, l'objectivité de la correction permet à l'étudiant d'obtenir lui-même ces informations et donc de s'impliquer dans sa propre formation (auto régulation).

Grâce à la concision du code, le professeur recueille (instantanément) un grand nombre de mesures (une par étudiant) et peut prendre des décisions rapidement, soit par l'ensemble de la classe, soit pour certains individus. Cette concision permet un encodage informatique rapide, ce qui favorise la prise en charge par l'ordinateur de traitements approfondis et la communication de conseils très détaillés.

3. La rançon des avantages offerts par les QCM

a) Le travail de mise au point d'une QCM

Pour concevoir une QCM, l'enseignant utilise les erreurs fréquentes ou graves qu'il a observées chez ses élèves. Parfois, il dispose aussi de listes d'erreurs établies systématiquement. Dans les deux cas, il profite d'un travail préalable considérable.

Mais il arrive que parcelles ressources n'existent pas. On découvre alors concrètement le prix à payer pour les obtenir.

La construction de QCM appelle une réelle compétence sur le fond et une riche expérience pédagogique.

La première rédaction d'une QCM doit faire l'objet d'un essai expérimental au cours duquel les réponses sont analysées selon des modèles, des formules et des procédures rigoureuses. En cas de résultats insatisfaisants, la QCM est modifiée et soumise de nouveau à l'expérimentation. Lorsqu'elle a enfin fait la preuve de sa clarté et de sa précision, la QCM peut être utilisée à des fins d'évaluation.

En cas d'examen externe (ou de concours), il est difficile pour des raisons de secret, de « prétester » les nouvelles QCM, ce qui est regrettable.

b) L'impossibilité d'une présentation orale

En principe, une QCM comporte une amorce dense (chaque mot compte) et des solutions proposées qui ne se différencient parfois que par un élément ou deux. L'étudiant doit donc pouvoir analyser ces données tout à son aise. De plus, les codes chiffrés sont purement arbitraires et il serait absurde de demander à l'étudiant de les mémoriser. Pour ces deux raisons, la présentation orale d'une QCM est, sauf cas exceptionnels, exclue.

Une telle présentation sera à la rigueur acceptée lorsque les solutions proposées sont extrêmement courtes et bien connues des élèves (ex. : « Un, deux, trois, quatre », ou « C'est un fait, c'est une opinion », ou « Masculin, féminin, neutre »). Quand des solutions sont susceptibles de se représenter souvent, le professeur peut les afficher en permanence avec un code.

Voici quelques affiches de ce type :

1. Nominatif.	1. Neutre.	1. Aigu.	1. Positif.
2. Vocatif.	2. Masculin.	2. Obtus.	2. Négatif.
3. Accusatif.	3. Féminin.	3. Droit.	3. Nul.
4. Génitif.			4. Indéterminé.
5. Datif.			
6. Ablatif.			
(Latin)	(Allemand)	(Géométrie)	(Algèbre)

Ce système, suggéré par Fromm (1970), constitue un compromis, puisque les solutions sont présentées sous forme écrite et l'amorce sous forme orale. Cette dernière peut alors être improvisée en cours de leçon. Mais, en dehors de pareils cas, l'improvisation est, en principe, incompatible avec la création des QCM.

La phase de mise au point est donc lourde. Cet investissement humain est-il raisonnable? Remarquons tout d'abord que tout instrument d'évaluation, quelle que soit sa forme, n'offre des garanties de fidélité et de validité qu'au prix d'un effort comparable. En contrepartie, un outil sûr peut être réutilisé. Quand une provision suffisante de questions existe, il suffit de choisir celles qui répondent aux nécessités du moment et à monter l'épreuve voulue.

La collaboration entre maîtres, indispensable aujourd'hui, comprend tout naturellement l'échange des instruments d'évaluation. Une solution d'avenir consiste à rassembler les QCM validées dans une banque de questions accessible aux utilisateurs potentiels.

c) La tentation de poser des questions de détail

Noizet et Caverni (1978, p. 155) rappellent que les questions doivent être « décidables », c'est-à-dire ne pas laisser de place à l'opinion.

« Cette condition de décidabilité est capitale puisque la réponse sollicitée est de nature « vrai » ou « faux ». Ainsi, par exemple, la question de savoir si tel écrivain a écrit ou non telle œuvre constitue un item décidable dans la mesure où la paternité d'une œuvre est susceptible d'être attribuée sans erreur. Mais la question de savoir si en dernière analyse tel écrivain peut être considéré comme un romantique constituerait un item indécidable ».

Les mêmes auteurs poursuivent :

« C'est sans doute la nécessité de présenter des items décidables qui conduit à choisir de faire porter les questions sur des points de détail. Cette tendance est fréquente – trop fréquente sans doute – et traduit une attitude de facilité. C'est elle qui explique le reproche, souvent adressé aux QCM, de conduire à une *parcelisation des connaissances*. Il faut considérer qu'il y a là un risque réel et ce risque est d'autant plus grave que, dans la perspective d'un examen par QCM, les candidats seraient induits à des stratégies de préparation en accord avec cette parcelisation.

Le souci du constructeur de QCM de se mettre, par une parcelisation du savoir, à l'abri des items indécidables a pour conséquence indirecte que les QCM font souvent appel à la mémoire beaucoup plus qu'à la logique. Il s'en faut cependant que cette conséquence soit inévitable. C'est ainsi que

des questions du type « la notion A se définit strictement comme ... » obligent à distinguer les conditions suffisantes à la définition de la notion des conditions simplement nécessaires. Cela suppose de la part du candidat un traitement de ses connaissances qui ne peut se réduire à une simple activité de mémoire ».

Nous tâcherons de montrer que la variété des consignes possibles permet de dépasser ce problème. Par ailleurs, le danger du recours à des solutions de facilité diminue à mesure que les professeurs maîtrisent mieux les techniques de sollicitation de la réponse.

d) La facilité de copier les réponses du voisin

Les deux étudiants ci-dessous ont convenu, avant d'entrer dans la salle d'examen, de communiquer par gestes, selon les codes suivants : mon nez = 1, mon œil = 2, mes cheveux = 3, mon oreille = 4.



Pour empêcher la fraude, on recourt au « brouillage » des numéros des questions et/ou des numéros des solutions proposées. Tous les étudiants reçoivent les mêmes questions et les mêmes solutions proposées, mais distribuées différemment dans des « formes parallèles » du même test. Souvent, on confie à l'ordinateur la remise de toutes les réponses en un seul ordre, et ce avant d'entamer la correction, identique pour tous. Il existe aussi des méthodes manuelles de « débrouillage ».

e) La possibilité offerte aux étudiants de fournir des réponses correctes en répondant au hasard.

Ces « choix heureux par ignorance », comme les appelle D'HAINAUT (1974), peuvent difficilement être distingués des « choix corrects par compétence ». A l'opposé, les choix incorrects sont, eux aussi, difficiles à interpréter : choix aléatoire ou erreur de « bonne foi » ?

Deux procédés ont longtemps tenté de compenser cet inconvénient : augmenter le nombre de solutions proposées et pénaliser les erreurs.

Jusqu'à ces derniers temps, la règle d'or était de construire des QCM avec au moins quatre solutions proposées, afin de réduire la part du hasard au maximum à 25%. Cependant ce principe est parfois inapplicable, comme pour la question suivante :

Quel est le genre du mot « Stein » en allemand ?
Masculin ? Féminin ? Neutre ?

On ne voit pas bien quelle quatrième solution réellement attractive pourrait être ajoutée. En pareil cas, le nombre de solutions s'impose de lui-même.

Pour dissuader les élèves d'effectuer des choix aveugles, on a, depuis le début de l'utilisation des QCM, utilisé la consigne : « Ne devinez pas aveuglément, car des points vous seront retirés en cas d'erreur ». Le retrait de points est connu sous le nom de *correction for guessing*, expression que de Landsheere (1979) propose de traduire par « correction pour divination ».

De nombreuses procédures de correction ont été utilisées, et une littérature abondante (surtout en anglais) existe sur ce sujet.

4. Trois désavantages fondamentaux des QCM

Les désavantages qui viennent d'être évoqués sont relativement secondaires ou évitables. Il n'en va pas de même pour les suivants :

a) L'impuissance des QCM à mesurer certains types de performances

C'est le cas, entre autres, pour l'expression spontanée, l'aptitude à rédiger et à exprimer sa pensée, l'invention de solutions nouvelles et même ... la reproduction de mémoire sans support.

Avec bien d'autres auteurs, Vandeveldde (1971) souligne la complémentarité des réponses « ouvertes » et des QCM :

C'est notre manque de sens critique qui nous fait reconnaître à la question de « création » des qualités qu'elle n'a pas.

Sans doute, « ce qui se conçoit bien s'énonce-t-il clairement », encore ne peut-on pas confondre comprendre et savoir. L'école se livre à des extrapolations abusives au travers de la connaissance. La connaissance étant la condition sine qua non de réussite, elle constitue un écran à l'égard de ce que nous voulons mesurer. L'individu examiné peut ne pas posséder la

matière sur laquelle porte l'épreuve, n'avoir pas dépensé l'énergie nécessaire à sa mémorisation. Ceci peut bien entendu lui être reproché, mais le grief ne porte dès lors plus sur son inaptitude à comprendre. Nous nous trouvons dans l'impossibilité de conclure à cet égard.

Il en va de même en analyse littéraire, notamment, qui se juge exclusivement sur la production et même sur la formulation de cette production. Nous jugeons, ce faisant, une performance parasitée par une variable dont il faudrait décider explicitement et a priori si on accepte de lui reconnaître l'importance que nous lui accordons inconsciemment.

C'est pour avoir trop souvent réuni en une même technique d'investigation une gamme importante de facteurs très différents (connaissance, compréhension, application, analyse, etc.) que l'image que nous nous faisons de l'élève est si imprécise.

Nos examens, nos épreuves de contrôle scolaire gagneraient beaucoup à se présenter de manière analytique et davantage sur le plan des activités mentales que sur le plan des matières. La question à choix multiple est susceptible de fournir une information précieuse dans cette perspective.

Les QCM ne sont qu'un outil parmi d'autres : il importe de recourir au mode d'évaluation le plus adéquat à chaque situation. Les QCM conviennent moins bien pour les performances complexes (réponses longues où la structure et l'expression jouent un grand rôle) que pour les performances isolables.

Dans le même esprit, Dudley (1973), qui travaille dans le domaine médical, reproche aux QCM de ne pas permettre de mesurer tous les aspects de la connaissance. Wood (1977) répond, toutefois, qu'un tel reproche revient à se plaindre du stéthoscope sous prétexte qu'il est de peu d'utilité pour l'examen des yeux et des oreilles...

b) La présentation de solutions erronées aux étudiants

On peut craindre que des erreurs ne se fixent dans la mémoire des étudiants. Ce danger paraît réel dans certains domaines. Il est, par exemple, contre-indiqué de présenter des questions d'orthographe d'usage sous forme de QCM. Mais d'autres domaines ne présentent pas la même vulnérabilité, parce qu'ils font davantage appel au raisonnement (ex. : l'orthographe grammaticale). Ce problème est discuté dans la section D ci-après.

c) L'incapacité des QCM à entraîner à la formulation de réponses

Nous ne pensons pas que l'exposition de l'étudiant à des QCM détériore la capacité de l'étudiant à rédiger sa réponse. Néanmoins, on peut objecter que toute question posée par écrit offre à l'étudiant une occasion de formu-

ler sa pensée, occasion que le recours aux QCM fait perdre. Il est évident que, aux niveaux scolaires où cet objectif est important, on veillera à ménager à l'étudiant un nombre suffisant d'opportunités pour s'exercer à formuler sa pensée.

d) La contraction du champ cognitif

On pourrait craindre que les QCM contribuent à simplifier la « vision du monde » des étudiants, à la réduire à un schéma, et par conséquent, à uniformiser les individus. Ce problème est discuté dans la section E ci-après.

D. Le problème de la mémorisation des solutions fausses

1. La critique de Skinner

« Toute solution fautive, dans un test à choix multiple, augmente la probabilité qu'un étudiant extraie un jour de sa mémoire défaillante la réponse incorrecte au lieu de la réponse correcte ».

Cette critique de Skinner (1961), souvent citée¹, appelle trois remarques.

Tout d'abord, les QCM ne sont pas les seules à présenter des solutions fausses aux étudiants. Il en va ainsi dans la vie de tous les jours.

Ensuite, on doit distinguer deux types d'erreur, comme le suggère Houziaux (1972, p. 92) :

« Il y a lieu de faire une distinction entre erreur relative et erreur absolue. Là où c'est avant tout le raisonnement qui est sollicité - c'est-à-dire, essentiellement, dans la résolution de problèmes - la réponse fautive constitue le résultat d'une opération mentale incorrecte et ne doit être rejetée que par rapport à un contexte, qui est l'énoncé du problème : erreur relative, qu'il n'y a pas d'inconvénient de proposer à l'élève. Par contre, dans les matières

¹ Skinner, 1968 p. 44, 45, 254.

qui font essentiellement appel à la mémoire, l'apprentissage se ramène à un conditionnement dont il est, en principe, préférable d'exclure l'erreur, celle-ci étant toujours absolue ; c'est le cas, par exemple, lorsque l'on veut faire mémoriser un terme technique ou enseigner l'orthographe d'usage. »

Enfin, même en cas d'erreur absolue, la critique de Skinner doit être nuancée selon la situation de testing, le contenu, le niveau taxonomique, les prérequis, les modalités d'application des QCM, le mode d'apprentissage de la matière, etc.

Ainsi, Eklund (1968) estime que le recours aux QCM en début d'apprentissage peut présenter un grand risque ; ultérieurement ce n'est plus le cas.

De même Noizet et Caverni (1978, p. 157), notent, à propos de l'effet structurant ou déstructurant des distracteurs sur le processus d'apprentissage :

« Il est vraisemblable que cet effet dépend des indices de détection des distracteurs qui ont été ou non fournis lors de l'enseignement, l'attention des élèves ayant été attirée ou non sur les erreurs ou les confusions susceptibles d'être commises ».

2. Les travaux de Preston

Le bien fondé de la critique de Skinner a été testé par Preston (1965).

Il a demandé à 79 étudiants (de fin d'enseignement secondaire) de définir dix mots de l'histoire des U.S.A., tels que sécession, embargo, etc.

- Lors d'un premier test (A), les étudiants devaient rédiger leur définition (questions ouvertes).
- Le second test (B) était à choix multiple.
- Le troisième test (C), dit de contrôle, se présentait de nouveau sous forme de questions ouvertes.

Preston recherche quatre séquences qu'il juge révélatrices des effets étudiés. Les deux premières sont destinées à tester les effets d'un choix incorrect lors du test à choix multiple. Les deux suivantes sont destinées à tester les effets d'un choix correct au test à choix multiple sur les réponses ultérieures à la question sous forme ouverte.

Les séquences se présentent comme suit :

Séquence 1 : Le distracteur choisi est sans rapport avec la réponse fournie en A, mais a inspiré la réponse en C.
Hypothèse : effet négatif d'un choix incorrect (cfr Skinner).

Séquence 2 : Le distracteur choisi est sans rapport avec la réponse fournie en A, mais n'a pas inspiré la réponse en C.
Hypothèse : effet nul d'un choix incorrect.

Séquence 3 : La réponse en A est incorrecte, mais la réponse choisie en B est correcte et a inspiré la réponse en C.
Hypothèse : effet positif du choix correct.

Séquence 4 : La réponse en A est incorrecte et la réponse choisie en B est correcte, mais n'a pas inspiré la réponse en C.
Hypothèse : effet nul d'un choix correct.

Les données montrent – quoique de façon variable selon les mots inducteurs – un effet négatif du choix incorrect et un effet positif du choix correct.

Preston a par ailleurs calculé, pour chaque individu, une tendance à suivre la séquence 1 (conforme à l'hypothèse de Skinner). Cet indice de tendance (appelé X_1) est obtenu par la formule :

$$X_1 = N_1 / (N_1 + N_2) \text{ où } N_1 = \text{nombre de séquences 1,} \\ N_2 = \text{nombre de séquences 2.}$$

Il a aussi calculé la tendance individuelle à suivre la séquence 3, opposée à l'hypothèse de Skinner par la formule :

$$X_3 = N_3 / (N_3 + N_4)$$

- Il a constaté les corrélations suivantes :
- entre le Quotient Intellectuel (QI) et X_1 : 0,33
 - entre le Quotient Intellectuel et X_3 : 0,30

On peut regretter le caractère artificiel de cette étude dans laquelle l'étudiant n'est pas informé de la réponse correcte après avoir effectué son choix. Ce biais, à lui seul, diminue considérablement la portée des résultats de cette ingénieuse expérience de Preston, résultats favorables à l'hypothèse de Skinner.

3. Les travaux de Karraker

D'autres auteurs ont essayé d'apporter les données expérimentales susceptibles de faire avancer le débat. C'est le problème de la pertinence des données, et surtout des hypothèses, qui va surtout être évoqué à partir de l'expérience réalisée par Karraker (1967).

L'auteur estime que la connaissance des résultats est le meilleur remède au danger sur lequel Skinner (1958, p. 176) attire l'attention en ces termes : « Les questions à choix multiple doivent contenir des réponses incorrectes plausibles ; or, c'est inadéquat dans le processus délicat de « façonnage » du comportement, car elles renforcent des formes incorrectes ».

Kaes et Zeaman (1960) ont observé que le fait d'informer l'étudiant que la réponse était incorrecte a un effet défavorable sur l'apprentissage. Ils ont proposé un test fait de QCM, en ont communiqué les résultats, réponse par réponse (correcte/incorrecte), puis leur ont administré à nouveau le même test. Ils ont constaté que, le plus souvent, les étudiants reviennent à leur première réponse fautive.

A propos de cette expérience, Karraker conclut que les étudiants ont appris la réponse incorrecte parce qu'ils y ont été exposés. Une telle conclusion paraît hâtive. Le phénomène pourrait être dû à une persévérance dans l'erreur : l'étudiant a une tendance à maintenir sa réponse, car elle est le résultat d'une erreur antérieure à la présentation de la QCM (comme le montrent les travaux de Powell et Isbister, 1974). Heureusement, l'expérience de Kaes et Zeaman montre aussi que 53 % des réponses incorrectes lors du test ont été remplacées par des réponses correctes lors du second passage à l'épreuve.

L'expérience de Karraker est, on va le voir, proche, dans son principe, de celle de Preston.

Un test de quarante questions a été présenté à septante-deux étudiants de candidature, répartis en trois groupes, selon les modalités expérimentales suivantes :

	Groupe 1 « Est informé des résultats »	Groupe 2 « N'est pas informé des résultats »	Groupe de Contrôle 3
SEANCE 1	Test QCM à trois solutions proposées pour chaque question		
SEANCE 2	Communication des réponses correctes	-	-
SEANCE 3	Test à réponse construite	Test à réponse construite	Test à réponse construite
Réponses correctes	M = 34 (S = 3)	M = 31 (S = 5)	M = 28 (S = 7)
Réponses incorrectes correspondant à un distracteur du test QCM	M = 3,54 (S = 2,53)	M = 6,33 (S = 2,82)	M = 3,70 (S = 2,51)

M = moyenne

S = écart type (ou sigma).

Les moyennes de réponses incorrectes correspondant à un distracteur du test ne diffèrent pas significativement entre le groupe 1 et le groupe 3, ce qui fait dire à Karraker que la lecture des distracteurs (lors de la séance 1) n'a pas eu d'effet.

La différence entre les moyennes des réponses correctes des groupes 1 et 3 est significative. La moyenne des réponses incorrectes du groupe 2 est différente de la moyenne des groupes 1 et 3, mais pas significativement.

Karraker signale que la création de son groupe 2 (identique au groupe expérimental constitué par Preston) n'a pas beaucoup de sens : dans la pratique pédagogique on donne connaissance des résultats, car les enseignants sont maintenant familiers avec les apports de l'analyse expérimentale du comportement et donc avec les lois du conditionnement opérant. Dans ce cadre, de nombreuses recherches ont montré que la connaissance immédiate des résultats améliore l'apprentissage scolaire (Berglund, 1969 ; Zontine, Richards et Rust, 1972 ; Betz et Weiss, 1976a, 1976b ; Strang et Rust, 1973 ; Beeson, 1973).

L'expérience de Karraker montre donc que l'effet redouté par Skinner se produit quand on crée des conditions anti-éducatives, mais qu'il n'y a pas lieu de le craindre dans des conditions pédagogiques normales. Au contraire, la supériorité de la performance du groupe 1 par rapport à celle du groupe 3 plaide en faveur de l'utilisation de QCM dans une pratique scolaire saine.

4. Discussion

On trouve dans la littérature expérimentale de nombreux comptes rendus d'expériences voisines de celles de Preston et Karraker.

Nous avons déjà dit combien les résultats varient avec les contenus. Nous venons de voir l'importance des apprentissages. Il faut aussi prendre en considération les instruments de mesure.

Il est probable, en effet, que la lecture des distracteurs a, dans bon nombre de cas, des effets subtils : le comportement des sujets n'est pas sensiblement changé, mais le trouble, le doute, s'insinue néanmoins dans leur esprit. C'est une constatation que chacun peut faire dans le domaine de l'orthographe d'usage au moyen de l'expérience que voici :

- Dictier des mots et recueillir les réponses (mots écrits) ainsi que des indices de certitude par lesquels les étudiants expriment dans quelle mesure ils sont sûrs de l'exactitude de la graphie de chaque mot.
- Présenter les graphies correctes.

c) Redicter les mêmes mots (recueillir à nouveau les indices de certitude).

Lorsque la graphie fournie par l'élève était déjà correcte, lors de la première réponse, il est probable qu'elle ne changera pas lors de la seconde. Néanmoins, on peut observer une augmentation du degré de certitude.

Une autre expérience consisterait à ne pas présenter en b les graphies correctes (communication des graphies correctes). La confiance dans les réponses resterait alors stationnaire lors de la phase c.

La recherche devrait s'attacher à déterminer l'intensité (à mesurer par des indices de certitude) et les variations selon les conditions (de contenu, de consigne, de niveau taxonomique, etc.) avec lesquelles ce phénomène se produit. Il importe de se concentrer sur des situations proches des conditions scolaires habituelles.

E. Le problème de l'uniformisation des individus et des sociétés

On prétend souvent que les QCM classiques reposent sur l'idée qu'à toute question il existe une réponse unique, et que leurs partisans font de l'apprentissage de cette réponse un idéal pédagogique, avec pour conséquence une uniformisation des individus.

Que les individus diffèrent moins les uns des autres à *certain*s égards, notamment intellectuels, n'a rien pour déplaire, pour autant que cette uniformisation se fasse par le haut. C'est tout le principe de la pédagogie de la maîtrise. Quoi de plus souhaitable que de voir le plus grand nombre possible d'individus atteindre l'excellence (au niveau scolaire considéré) dans des disciplines telles que la grammaire, la géographie, les mathématiques, etc ?

Mais n'est-il pas des domaines où c'est la variété même des comportements qui constitue la richesse ? On cite souvent à ce propos le cas de l'expression verbale. C'est pourquoi les travaux de Kendler retiendront notre attention.

1. Les travaux de Kendler

Kendler (1959), à propos du vocabulaire (anglais), rapporte une expérience de Jenkins et Russell (1958) qui ont utilisé une même épreuve d'association de mots (épreuve de cent mots de Kent-Rosanoff) auprès d'étudiants de l'Université du Minnesota, et cela deux fois en vingt-trois ans d'intervalle.

Dans ce genre de test, l'expérimentateur prononce un mot et les sujets écrivent le mot que cela leur suggère. Pour chaque mot dicté, on peut rechercher le mot-réponse le plus populaire (celui qui fut le plus souvent écrit) et calculer son pourcentage d'utilisation sur l'ensemble des réponses des sujets. En moyenne, pour l'ensemble des mots, cet indice atteignait 29 % en 1929 et 38 % en 1952. Pour Kendler, ces résultats semblent indiquer que la société américaine devient de plus en plus homogène dans ses habitudes verbales.

Il serait pour le moins aventureux de considérer cette conclusion comme définitive. Les populations considérées ne sont peut être pas comparables : en 1929, un pourcentage beaucoup plus faible de personnes fréquentaient l'Université qu'en 1952.

Kendler soulève lui-même cette objection et ajoute qu'à l'inverse de l'anglais, les associations de mots chez les Francophones et les Germanophones apparaissent moins uniformes. Bien que conscient de diverses explications possibles (l'influence grandissante des médias sur la vie de tous les jours, l'uniformisation du système scolaire), Kendler n'en considère pas moins la situation comme très préoccupante : « S'il se confirme que l'originalité dans les associations de mots facilite la résolution de problèmes, comme certains auteurs¹ le prétendent, nous sommes confrontés à une menace de notre créativité scientifique nationale ».

Kendler justifie ses craintes en observant que les découvertes scientifiques faites depuis 1940 aux U.S.A. sont dues, en majorité, à des Européens récemment émigrés.

Si les appréhensions et les explications de Kendler nous paraissent exagérées, il n'en reste pas moins que la vigilance s'impose. Nous pensons avoir suffisamment précisé les limites des QCM et avoir assez mis en garde contre leur utilisation abusive (par seule commodité, par exemple).

2. Où sont les vrais problèmes ?

Une polémique du même genre avait été engagée il y a plus de dix ans à propos de l'enseignement programmé linéaire. On reprochait à ce dernier de transformer les étudiants en robots, de leur imposer une voie rigide d'apprentissage, même si le rythme était individuel. On agissait aussi le spectre d'une école où les étudiants n'auraient plus de contacts avec leurs professeurs.

¹ Guilford (1967) et Amitzman *et al.* (1958).

En fait, que s'est-il passé ? L'enseignement programmé a pris sa place parmi les divers moyens d'enseignement de façon si modeste qu'un certain nombre de professeurs n'ont pas encore eu l'occasion de s'en servir. Actuellement, les maîtres restent juges de l'opportunité d'y recourir ou non. Ces cours sont présentés aux étudiants comme des occasions d'apprendre et non comme des méthodes imposées.

Mais il ne suffit pas de veiller à éviter les emplois abusifs ou néfastes d'une méthode, il faut développer d'autres méthodes complémentaires pour la formation de l'individu.

Par exemple, on peut très bien continuer à produire des cours programmés linéaires, tout en construisant des systèmes ramifiés² et des séquences d'entraînement à la résolution de problèmes (S.E.R.P.).

Alors que les cours programmés vont du simple au complexe, les S.E.R.P. procèdent de la façon inverse : l'étudiant est d'abord confronté à un problème dans toute sa complexité. Il est amené à l'analyser progressivement en composantes plus simples³.

L'enseignement programmé et ces séquences sont des modalités d'apprentissage qui se complètent et qui contribuent, avec bien d'autres méthodes, à une formation harmonieuse, sans excès mais aussi sans carences.

Il en va de même pour les QCM : technique parmi d'autres, elle ne peut s'imposer au détriment d'autres pratiques. En particulier, il importe d'accorder une importance spéciale à l'évaluation des performances à correction complexe. Le livre de De Bal, De Landsheere et Beckers (1977) situe bien le problème et décrit des techniques appropriées à ce cas.

F. Synthèse : Régulations, évaluations et QCM

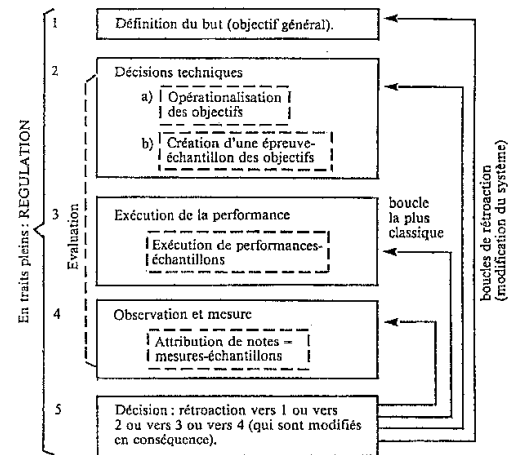
1. Régulation et évaluation pédagogique

La QCM n'est qu'une technique de l'évaluation pédagogique qui elle-même n'est pas une fin en soi, mais fait partie d'un mécanisme de régulation au service d'un objectif. Dans un processus régulé, la performance

² Pacelbro : Programme autocorrectif à embranchements sous forme de livre brouillé à réponse ouvertes (décrit dans Leclercq, Donnay, De Bal, 1977).

³ D. Leclercq (1982).

observée est comparée à une performance de référence (« recherchée » ou « attendue ») et est modifiée quand elle s'écarte du seuil déterminé. Le processus de régulation peut être décomposé en cinq opérations (Leclercq 1976).



On voit sur ce schéma que l'évaluation pédagogique coïncide avec trois des opérations du processus de régulation, c'est pourquoi on la confond si souvent avec le processus entier.

Ce schéma de la régulation nous sera utile pour situer les avantages et inconvénients des QCM déjà présentés.

3. Conclusions

Une QCM est une situation de choix particulière, définie par ses composantes et par le fait que la solution correcte, connue au préalable de l'enseignant, est invariable, quel que soit l'individu qui répond à la question.

Les QCM ne peuvent résoudre tous les problèmes de l'évaluation pédagogique. Il s'agit d'une technique parmi d'autres, qui doit être choisie en fonction de son adéquation au contenu, au niveau de comportement sollicité, aux contraintes du moment.

Les avantages et les inconvénients des QCM sont réels. On s'ingéniera à exploiter les premiers et à réduire, voire supprimer, les seconds.

CHAPITRE 2

Les consignes classiques

- A. La structure profonde d'une QCM
 1. Question et sous-questions
 2. Le simplexe des réponses
- B. Les questions vrai-faux
 1. Variables affectant leur difficulté
 2. Compensations de la tendance à l'acquiescement
 3. Questions VRAI-FAUX et fidélité des mesures.
- C. Une taxonomie des consignes
 1. Codes Q
 2. Codes R
- D. Les consignes classiques
- E. Variantes des consignes classiques
 1. « La meilleure » solution
 2. La solution « Aucune »
 3. La solution « Toutes »
 4. La consigne ABDO
 5. Les associations composées
 6. Les QCM « A, B, Relation »
- F. La consigne « appariement »
 1. Correspondance biunivoque
 2. Asymétrie par excès de solutions
 3. Asymétrie par défaut de solutions
 4. Asymétrie par défaut et par excès de solutions
- G. La consigne « association à terme exclu »
- H. La consigne « questions solidaires »
 1. Les QCM en cascade.
 2. Les QCM partiellement solidaires.
 3. La liaison des questions lors de la correction.
- I. Conclusions.

Une expérience menée sur six cent cinquante-six étudiants en médecine devant répondre à seize questions a permis à ces auteurs (p. 172) d'observer que la performance varie du simple au double d'une condition expérimentale à l'autre (27 % de patrons exacts dans le premier cas contre 53 % dans le second cas).

b) L'influence de la formulation (positive ou négative)

Si l'on accepte l'hypothèse que le savoir est constitué essentiellement d'énoncés vrais, on peut se demander si ces énoncés sont toujours affirmatifs, ou s'il arrive que l'énoncé-signal soit négatif¹.

Pour Fabre et Noizet (1977), de tels énoncés signaux négatifs existent, mais en minorité, quand l'attention des élèves a été attirée sur certaines confusions à ne pas commettre, sur certaines assimilations à ne pas opérer².

La plupart des énoncés signaux étant affirmatifs, les savoirs seraient organisés comme un réseau de relations positives. Les solutions d'une QCM portant sur ces énoncés signaux affirmatifs sont vraies quand elles sont à la forme positive et fausses quand elles sont à la forme négative. Selon Fabre et Noizet (1977), c'est ce qui explique que ces deux types de solutions sont bien mieux réussies que les vraies à la forme négative et les fausses à la forme affirmative. Les pourcentages de réussite moyens observés sur septante-huit sujets devant répondre par « jugement multiple » à seize questions de cinq solutions chacune se présentent en effet comme suit :

Formulation	Solution		
	Vraie	Fausse	
Affirmative	83,3	50,7	67,0
Négative	52,3	76,7	64,5
	67,8	63,7	

¹ De tels « énoncés » correspondent aux principes de Gagne (1965). Par exemple : « Les oiseaux volent » (énoncé vrai positif), « Les poules n'ont pas de dents » (énoncé vrai négatif).

² C'est ce qui est appelé « discriminations entre unités d'apprentissage », dans la construction d'un cours programmé (voir Leclercq, Donnay, et De Bal, 1977, p.108-111) ; c'est l'objet principal de la technique dite « matrice de Davies ».

· Selon ces mêmes auteurs, la légère dissymétrie en faveur des « vraies négatives » par rapport aux « fausses négatives » viendrait de l'existence des quelques énoncés-signaux négatifs dans la connaissance. Mais les « vraies-affirmatives » (83,3) l'emportent sur les « fausses-négatives » (76,7). Cette constatation confirme une hypothèse maintes fois exprimée, à savoir qu'il existerait une *tendance générale à l'acquiescement* (répondre affirmativement en cas de doute).

c) L'influence du caractère VRAI ou FAUX d'une solution sur son attrait

Bandenburg et Whitney (1972), ont montré que les questions dont la réponse correcte est « VRAI » paraissent plus claires, c'est-à-dire moins ambiguës que les questions dont la réponse correcte est « FAUX ». Dans leur expérience, la facilité moyenne est de 75 % pour les premières et de 65 % pour les secondes.

D'autres biais de construction peuvent aussi être invoqués. Ainsi, après avoir examiné de nombreuses questions VRAI-FAUX, Remmers et Gage (1955, p. 89) signalent les tendances suivantes. Les propositions contenant les termes *seulement, tous, aucun, rien, jamais, toujours*, ont souvent FAUX pour réponse correcte.

Par contre, les propositions contenant les termes *doit, devrait, peut, quelque, quelque fois, la plupart, souvent, généralement*, ont souvent VRAI pour réponse correcte.

Noizet et Caverni (1978) commentent l'une de leurs expériences menée en collaboration avec Fabre (1975) :

« Elle apporte des informations intéressantes sur les erreurs commises par les candidats. Ces erreurs sont de deux types : prendre le faux pour le vrai (SQF donnant naissance à une SRV) ou prendre le vrai pour le faux (SQV donnant naissance à une SRF). L'erreur de type 1 correspond à ce qu'on appelle fausse alerte... l'erreur de type 2 à ce qu'on appelle une omission (non détection d'un signal (Noizet et Fabre, 1975). Or, les résultats indiquent une dissymétrie dans les erreurs commises : les erreurs consistant à prendre le faux pour le vrai sont significativement plus nombreuses que celles consistant à prendre le vrai pour le faux (70,8 % vs 29,2 % pour les questions 2 vrai 3 faux et 3 vrai 2 faux).

Ce décalage très important, que confirment toutes les expériences menées sur des examens réels, est révélateur de la psychologie du candidat. Il est susceptible de deux explications que la théorie dite de la détection du signal (Green et Swets, 1966) nous a appris à distinguer. En termes de probabilité du signal, le candidat surestime le nombre de sous-questions

« Cette question peut être décrite comme composée de deux sous-questions (SQ) :

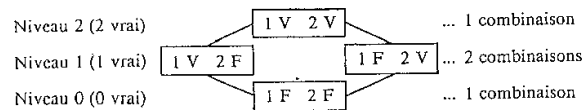
SQ 1. L'ataxie est associée à un syndrome cérébelleux.
VRAI-FAUX.

SQ 2. L'ataxie est associée à un syndrome tabétique.
VRAI-FAUX.

Toute solution à la QCM est donc la combinaison de n (ici 2) sous-réponses. Par exemple, la solution 1 est la combinaison de la sous-réponse : « La réponse à la SQ 1 est V » et de la sous-réponse : « La réponse à la SQ 2 est F » (p.41).

2. Le simplexe des réponses

Les deux auteurs précédents représentent toutes les combinaisons possibles des sous-réponses (V = vrai et F = faux) aux diverses sous-questions (1 et 2), dans un schéma appelé « simplexe » et comportant plusieurs niveaux :



1 V signifie : « La réponse correcte à la sous-question 1 est V ».
1 F signifie : « La réponse correcte à la sous-question 1 est F ».

Le niveau 0 correspond à l'absence de sous-réponse correcte (V).
Le niveau 1 correspond à la présence d'une sous-réponse correcte (V).
Le niveau 2 correspond à la présence de deux sous-réponses correctes (V).

Deux combinaisons reliées entre elles par un trait (ici oblique) ne se distinguent que par une seule sous-réponse.

On voit qu'en partant de n sous-questions, on obtient 2ⁿ combinaisons (ici 4).

Dans l'exemple ci-dessus, relatif à l'ataxie, toutes les combinaisons du simplexe ont été mises en œuvre dans la QCM. D'habitude, on n'utilise que les combinaisons de niveau 1 (une seule sous-réponse correcte), ce qui donne des QCM simples.

B. Les questions vrai-faux

Ebel (1969, 1970 et 1971) avance que la connaissance s'acquiert fondamentalement par le jugement de propositions VRAI-FAUX. C'est pour cette raison, entre autres, qu'il a toujours fortement recommandé le recours aux questions de ce type.

Noizet et Caverni (1978) font également l'hypothèse générale que « les savoirs s'organisent essentiellement sous forme d'énoncés vrais ».

Ces énoncés signaux sont susceptibles d'être facilement reconnus parmi les solutions d'une QCM, comme dans les énoncés des questions VRAI-FAUX.

Fabre et Noizet (1977) ont réalisé une série d'expériences qui éclairent cette hypothèse. Les résultats de l'une d'entre elles leur font croire qu'« il est toujours plus facile de repérer le vrai que le faux » (p. 177). D'autre part, ils arrivent aussi à la conclusion qu'« il est plus facile de repérer ce qui se détache » (faux parmi plusieurs vrais ou l'inverse) (p.177).

Ces deux effets iront dans le même sens ou en sens inverse, selon la façon dont la question se présente.

1. Variables affectant les résultats à une question VRAI-FAUX

a) L'influence de la consigne sur le processus

Deux types de consignes créent des situations très différentes quant à la reconnaissance des énoncés vrais ou « qui se détachent ».

Dans le premier type de consigne, l'étudiant doit juger *séparément* chacune des solutions proposées. C'est ce que Fabre et Noizet (1977) appellent le *jugement multiple*.

Dans le second type de consigne, l'étudiant est averti du nombre de solutions correctes (souvent une seule), et de solutions fausses. Fabre et Noizet (1977) diraient que l'étudiant est informé du nombre de sous-questions VRAIES et du nombre de sous-questions FAUSSES, et qu'il est dès lors averti de restrictions sévères sur le simplexe.

Noizet et Caverni (1978, p. 171) pensent que la *tâche d'identification* (repérage des énoncés conformes aux savoirs acquis ou aux énoncés signaux) suscitée par la première consigne se transforme en *tâche de discrimination* avec la deuxième consigne. L'étudiant a alors tendance à choisir un patron de réponse, c'est-à-dire un ensemble de sous-réponses, ce qui lui facilite la tâche (une fois les x solutions correctes repérées, les autres peuvent être considérées comme incorrectes).

Introduction

Dans le précédent chapitre, nous avons rappelé que l'on peut interroger un étudiant sur un même contenu en recourant à diverses techniques de sollicitation de la réponse : questions ouvertes, questions fermées, etc.

Les QCM elles-mêmes offrent un grand nombre de possibilités de questionnement. Dans le présent chapitre, nous examinerons à quels grands types de QCM on aboutit en faisant varier deux aspects seulement de la consigne :

- la description des questions (codes Q) ;
- la façon d'y répondre (codes R).

Le nombre de consignes possibles étant assez élevé, nous donnerons un code à chaque variation, que nous illustrerons par une QCM.

Avant d'aborder cette taxonomie, nous analyserons la structure profonde d'une QCM et les questions VRAI-FAUX.

A. La structure profonde d'une QCM

1 Question et sous-question

Noizet et Fabre (1975, p. 41) montrent qu'une QCM peut être décomposée en plusieurs alternatives (questions VRAI-FAUX). Ils donnent l'exemple suivant :

L'ataxie est associée à :

1. Un syndrome cérébelleux.
2. Un syndrome tabétique.
3. Tous les deux.
4. Aucun des deux.

et ils le commentent :

vraies et sous-estime celui des sous-questions fausses. En termes de coût de la réponse, le candidat considère qu'il est moins grave de déclarer vrai un énoncé faux que de déclarer faux un énoncé vrai.

Dans un cas comme dans l'autre, l'effet produit est une fréquence de sous-réponses « vrai » très supérieure à la fréquence attendue. Tout se passe par conséquent comme si, pour le sujet questionné, les énoncés se situaient sur une échelle de vraisemblance, du moins vraisemblable (énoncé assurément faux) au plus vraisemblable (énoncé assurément vrai), la continuité de l'échelle traduisant l'absence de dichotomie du vrai et du faux. Sur cette échelle, le sujet se fixe un seuil de décision et il répond « vrai » si l'énoncé se situe, sur l'échelle, au-delà de ce seuil, « faux » dans le cas inverse. Ce qu'indique le décalage entre les deux types d'erreurs, c'est que ce seuil est situé à un niveau de vraisemblance très bas. »

d) L'effet de la tournure négative de la solution proposée

Peterson et Peterson (1976) ont demandé à des étudiants de répondre à des questions portant sur un texte qu'ils venaient de lire. Chaque question consistait en une proposition (vraie ou fausse) rédigée affirmativement ou négativement.

Les négatives vraies amenèrent le plus d'erreurs, suivies par les négatives fausses, puis les affirmatives fausses.

Ce résultat porte à éviter les questions négatives. Notons cependant qu'il diffère légèrement des résultats de Wason (1961) quant à l'ordre de facilité des deux derniers types de propositions. L'expérience de Peterson et Peterson est critiquée de façon sévère par Wood (1977 ; p. 213).

Noizet et Caverni (1978, p. 160) notent que le traitement psycholinguistique des propositions négatives suscite des difficultés qui se marquent, par exemple, par des temps de vérification plus longs (Gough, 1965).

2. Compensations de la tendance à l'acquiescement

De nombreux auteurs¹ émettent l'hypothèse qu'il existe une tendance à l'acquiescement. Dans un document de synthèse, Brandenburg et Whitney (1972) discutent diverses méthodes destinées à compenser cette tendance à répondre affirmativement.

Considérons diverses méthodes possibles de calcul de points, ainsi que divers barèmes.

a) Absence totale de correction-pour-acquiescement

On accorde un point par réponse correcte et 0 par erreur ou omission.

¹ Cough et Keniston, 1960 ; Cronbach, 1942, Miklich et Gordon, 1968 ; Bentler, Jackson et Messick, 1971.

b) la correction pour divination classique

Les tarifs sont les suivants : + 1 point (réponse correcte), 0 (omission) et - 1 (erreur).

c) Le déséquilibre du rapport numérique entre réponses correctes et réponses incorrectes

La méthode la plus simple consiste à introduire dans le test plus de questions à réponse correcte FAUX que de questions à réponse correcte VRAI. Mais quelle doit être la proportion idéale ? Celle-ci doit probablement varier selon les individus.

d) Les questions pairées (correction n')

Larkins et Shaver (1968) ont décrit une autre méthode¹ destinée à neutraliser la tendance à l'acquiescement. On écrit deux formes « parallèles » de chaque question ; la forme A présente la réponse correcte VRAI et la forme B la réponse correcte FAUX. Ces auteurs proposent que le point ne soit accordé que si l'élève a répondu correctement aux deux questions (A et B). Les notations des formes A et B d'une même question (ou d'un même concept de base) sont donc solidaires. On peut accorder un point en cas de réussite, 0 en cas d'échec à l'une des deux formes et -1 en cas d'erreur aux deux formes.

e) Les questions pairées (correction n')

On accorde deux points quand les réponses aux deux questions d'une même paire sont correctes, 1 point en cas de succès seulement à la question B (dont la réponse correcte est FAUX) et 0 en cas de succès à A, mais pas à B.

f) Les questions pairées (correction n')

On accorde 4/3 de point en cas de succès à une question B (RC = FAUX) et 2/3 en cas de succès à une question A (RC = VRAI).

Brandenburg et Whitney (1972) ont comparé les procédures a, d, e et f ; ces deux dernières procédures ont été inventées pour les besoins d'une expérience. Leurs résultats montrent que les mesures obtenues avec les procédures d et e sont supérieures aux autres quant à la validité et à la fidélité. En revanche, leurs données ne montrent pas de tendance à l'acquiescement.

¹ Préconisée aussi par Ebel (1965).

g) Propositions pour une étude expérimentale

L'hypothèse de tendance à l'acquiescement ne peut être rejetée sur la base des seules expériences relatées ci-dessus. On pourrait tenter de la tester en recourant aux degrés de certitude. On pourrait aussi étudier la cohérence du comportement de l'étudiant. En effet, si un étudiant se trompe aux deux formes (A et B) d'une même question, il se montre cohérent: il ne se contredit pas. Les degrés de certitude pourrait peut-être jeter quelque lumière sur ce problème. On peut, en effet, s'attendre à ce que les réponses cohérentes soient exprimées avec une conviction plus grande que les réponses contradictoires. Dans une telle étude, aucune des procédures de notation exposées ci-dessus ne serait retenue: on utiliserait un barème de notation lié aux degrés de certitude.

3. Questions VRAI-FAUX et fidélité des mesures.

On sait qu'il faut moins de temps pour répondre à une QCM à 5 solutions proposées qu'à 5 questions VRAI-FAUX, mais il ne faut pas cinq fois moins de temps. On sait aussi que le nombre de solutions proposées par QCM accroît la fidélité de l'épreuve. La fidélité théorique peut être calculée en fonction du nombre de solutions proposées et du nombre de questions posées. (voir p. 144)

C. Une taxonomie des consignes

Voici les variantes que l'on peut introduire dans la consigne en ce qui concerne la description des questions et des réponses.

1. Codes Q: Les questions de l'épreuve

- Q0 - Une seule (ou aucune) des solutions proposées est correcte.
- Q1 - Une seule solution proposée est correcte.
- Q2 - C solutions proposées sont correctes.
- Q3 - Une ou plusieurs des solutions proposées est (sont) correcte(s).
- Q4 - Une, plusieurs ou aucune des solutions proposées est (sont) correcte(s).
- Q5 - La qualité des solutions proposées varie sur un continuum de V1 à V2.

2. Codes R: La façon de répondre

- R0 - Fournissez une réponse pour chaque solution proposée, l'abstention constituant une réponse.
- R1 - Fournissez une réponse (ou omettez).
- R2 - Fournissez r réponses (N.B. r est supérieur à 1 et est inférieur à k, le nombre de solutions proposées).
- R3 - Fournissez une ou plusieurs réponses (N.B. le nombre n'est pas précisé et peut-être égal à k, le nombre de solutions proposées).
- R4 - Fournissez une réponse pour chaque solution proposée (1 - VRAI - 2 - FAUX - 0 - OMISSION).

Ces six variantes pour Q et ces cinq variantes pour R déterminent trente consignes théoriquement possibles, dont les codes figurent dans le tableau suivant.

		Codes Q						
		0	1	2	3	4	5	
Codes R	0	00	10	20	30	40	50	(1)
	1	01	11	21	31	41	51	Consignes classiques
	2	02	12	22	32	42	52	
	3	03	13	23	33	43	53	
	4	04	14	24	34	44	54	

La consigne de loin la plus utilisée est la consigne QR=11: on annonce à l'étudiant qu'une seule solution présentée est correcte, et qu'il doit fournir une seule réponse.

Le présent tableau a précisément pour but d'aider à dépasser cette utilisation routinière des QCM.

Nous appelons « consignes classiques » celles où R=1.

Chaque consigne sera accompagnée d'un exemple.

¹ Voir Chapitre 4, section A1.

D. Les consignes classiques

Q0 : Une seule (ou aucune) des solutions proposées (n') est correcte.
R1 : Fournissez une seule réponse (ou omettez).

1) *Consigne QR = 01*

La capitale de l'Angleterre est :

1. Glasgow
2. Liverpool
3. Manchester
4. Dundee
5. Aucune des quatre.

Choisir la solution 5 (l'étudiant dit savoir) est tout différent de l'omission (l'étudiant dit ne pas savoir).

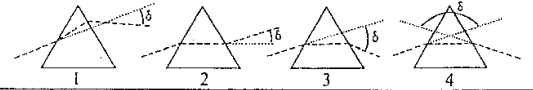
Nous discuterons ce point de manière plus approfondie dans la section E2 ci-après.

2) *Consigne QR = 11*

Q1 - Une seule solution proposée est correcte.

R1 - Fournissez une seule réponse (ou omettez).

Un rayon lumineux tombe sur un prisme d'indice $n = 1,5$ placé dans l'air. Quel schéma représente correctement l'angle de déviation δ ?



RC=3 (Question due à F. Debot).

3) *Consigne QR = 21*

Q2 - Deux solutions proposées sont correctes.

R1 - Fournissez une seule réponse (ou omettez).

Deux des formules suivantes permettent de calculer la circonférence du cercle. Choisissez l'une d'entre elles.

- | | | |
|--------------|----------------|----------------------|
| 1. $2\pi R$ | 3. $2\pi R^2$ | 5. $\frac{\pi R}{2}$ |
| 2. πR^2 | 4. $(\pi R)^2$ | 6. πD |

RC = 1 ou 6

4) *Consigne QR = 31*

Q3 - Une ou plusieurs solutions présentées sont correctes.

R1 - Fournissez une seule réponse (ou omettez).

Un patient se présente en état de choc au service des urgences. La mesure de sa pression artérielle donne les valeurs 8/6 ; la prise de son pouls indique une fréquence cardiaque de 120 ; l'électrocardiogramme montre des signes d'infractus ainsi que de nombreux extrasystoles ventriculaires. Mis en présence de ces éléments diagnostiques, vous initiez le traitement suivant :

1. une perfusion de noradrénaline.
2. une perfusion de métaraminol (aramine)
3. une perfusion de dopamine
4. une injection sous-cutanée d'atropine
5. une perfusion de lidocaïne
6. une perfusion de rhéonacrodex
7. une injection de forte dose (1g) de méthylprednisolon (solu-medrol)

RC = 3 ou 5 ou 7

Cette consigne s'indique lorsque plusieurs solutions sont correctes, mais qu'une réponse exclut les autres. C'est assez souvent le cas, dans le choix des moyens pour atteindre un but, et lorsqu'un seul moyen peut être employé.

5) *Consigne QR = 41*

Q4 - Une, plusieurs ou aucune des solutions proposées est (sont) correcte(s).

R1 - Fournissez une seule réponse (ou omettez).

Parmi les fruits suivants, le(s) quel(s) apparten(en)t au groupe botanique des POMMES ?

- | | |
|-----------------|------------|
| 1. Les poires. | 4. 1 et 3. |
| 2. Les pommes. | 5. 1 et 2. |
| 3. Les cerises. | 6. 2 et 3. |

RC = 5

Deux niveaux du simplexe des sous-questions ont été utilisés.

Rappelons (voir section A) qu'à partir des trois propositions de base (pouvant donner lieu à trois questions VRAI-FAUX), on ne peut construire que 8 solutions d'une QCM.

6) *Consigne QR = 51*

Q5 - La qualité des solutions proposées varie sur un continuum (de -4 à +3) avec une moyenne de -0,5.

R1 - Fournissez une seule réponse (ou omettez).

Quelle est la traduction qui convient le mieux pour : To be or not to be :

1. Etre ou pas
2. Etre ou non
3. Etre ou ne pas être
4. Pour être ou ne pas être
5. Etre ou bien ne pas être
6. Etre ou n'être pas.

Conséquences (tarifs)

- 2
- 1
- +3
- 4
- 0
- +1

(Question due à J. Noël)

A chaque solution, est attachée une valeur (inconnue de l'étudiant) d'autant plus positive qu'elle est jugée souhaitable par le correcteur, et d'autant plus négative qu'elle est jugée indésirable ou incorrecte. Les conséquences sont variables d'une question et d'une solution à l'autre. On doit annoncer à l'étudiant la valeur inférieure (V1), la valeur supérieure (V2) et la valeur moyenne (VM) de ces pondérations.

On constate que la somme des conséquences fixées pour cette question vaut - 3 points. La conséquence moyenne vaut donc -0,5 points, ce qui devrait décourager de répondre au hasard. De telles pondérations risquent d'aboutir, au total du test, à des scores nettement inférieurs aux résultats des évaluations classiques. Une telle sévérité se justifie par la gravité des erreurs ou par la facilité de la question. Dans l'exemple suivant, on constatera que la somme des conséquences est positive (+ 2 points) :

Consigne QR = 51

Q5 - La qualité des solutions proposées varie sur un continuum (de -3 à +4) avec une moyenne de +2.

R1 - Fournissez une seule réponse (ou omettez).

Quelle est la meilleure traduction de « That's the question » ?

1. C'est une question
2. C'est là la question
3. Voilà la question
4. Voici la question
5. C'est la question
6. Ça c'est la question
7. La question est là
8. C'est toute la question
9. C'est vraiment le problème

Conséquences

(inconnues de l'élève)

- 3
- +3
- +4
- 2
- +1
- 1
- +2
- 0
- 2

Exemple (dû à J. Noël)

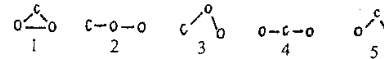
On voit que la procédure d'attribution des points est, dans ce cas, très éloignée des barèmes classiques fixés pour l'ensemble des questions d'un test. Les tarifs sont variables.

E. Variantes des consignes classiques

On rencontre beaucoup de formes de consignes aux noms originaux. Certaines comportent en effet des particularités qui en font des variantes réellement différentes. D'autres ne se distinguent des consignes déjà examinées que par des détails.

1. La consigne « choisissez la meilleure solution » (consigne QR = 11).

La forme de la molécule de dioxyde de carbone est le mieux représentée par le diagramme :



Il s'agit d'une variante intéressante de la consigne où QR = 11, à condition que l'on précise *en quoi* la solution doit être la meilleure. Il est, en principe, plus difficile de répondre à une QCM de la forme ci-dessus qu'à une QCM classique où la réponse correcte est opposée à des solutions incorrectes. En effet, dans la forme de QCM ci-dessus, plusieurs solutions peuvent être considérées comme acceptables (4 et 5), et il faut les classer entre elles pour, finalement, choisir la meilleure.

La structure profonde de ce type de question ne peut plus être exprimée par un simplexe et on ne peut la remplacer par plusieurs questions VRAI-FAUX.

2. La solution « aucune » (Consigne QR=11)

La solution « aucune » constitue une façon commode d'utiliser la combinaison de niveau 0 du simplexe. Cette solution doit alors figurer en dernière position et être précédée de combinaisons de niveau 1 ou de niveaux plus élevés.

Quand la solution « aucune » (niveau 0) figure à la suite des solutions d'une QCM, on laisse la porte ouverte à un grand nombre de sous-questions possibles (dans le sens de Noizet et Fabre), même si la QCM ne les évoque pas. C'est le cas pour la question qui suit :

Quel instrument symbolise le loup dans « Pierre et le loup » de Prokofiev ?

1. La flûte.
2. Le cor.
3. Le hautbois.
4. Aucun des trois.

Trois sous-questions sont suggérées par la QCM :

1. Le loup est représenté par la flûte. V - F
2. Le loup est représenté par le cor. V - F
3. Le loup est représenté par le hautbois. V - F

La présence de la quatrième solution (aucun) permet d'envisager d'autres sous-questions possibles (par le saxophone, par le basson, etc.).

Noizet et Fabre (1975) font remarquer que l'introduction de la solution « Aucun des trois » fait passer l'attractivité (théorique) de chacune des trois solutions de 1/3 à beaucoup moins de 1/4 (ici, à 1/10, si dix sous-questions sont envisagées). Dans un tel raisonnement, cette quatrième solution serait plus attractive que les autres puisqu'elle « rassemblerait » ici sept sous-questions sur dix.

Dudycha et Carpenter (1973) ainsi que Choppin (1974 a) ont montré expérimentalement que les questions recourant au rejet des solutions proposées sont plus difficiles que les autres.

Bishop, Knap et Mc Intyre (1969) constatent que la différence la plus importante entre les distributions de réponses à des QCM (consigne habituelle) et à des réponses ouvertes se manifeste pour les QCM contenant la solution « Aucune des précédentes ».

Williamson et Hopkins (1967) n'ont pas observé de modifications de la validité des épreuves, ni de leur fidélité, lorsque cette solution était employée. Ce résultat amène Choppin (1974 a) à conclure au peu d'efficacité de cette option. Ce jugement est, à notre avis, trop sévère. Le rejet de

toutes les solutions proposées est une formule qui se justifie surtout quand on veut amener l'étudiant à produire (intérieurement) sa réponse avant de consulter les solutions proposées.

3. La solution « toutes » (Consigne QR=31).

La solution proposée « toutes » correspond à la combinaison du niveau le plus élevé du simplexe. Elle doit alors figurer en dernier lieu et être précédée de combinaisons de niveau 1 ou de niveaux plus élevés, mais suivie de la solution « aucune » (combinaison de niveau 0).

La consigne générale doit préciser que, si cette solution est correcte, elle a la priorité sur toutes les autres qui le sont aussi (partiellement).

La présence de la solution « aucune » nous a fait créer un code Q particulier (Q=0), car l'étudiant était obligé d'envisager une réponse qui pourrait ne pas être sous ses yeux. Ce n'est pas le cas pour la solution « toutes », qui n'est qu'un aspect particulier des consignes où le code Q=3.

4. La consigne ABDO (Consigne QR=31).

D'Hainaut (1973) appelle Abdo les QCM comportant les quatre solutions suivantes :

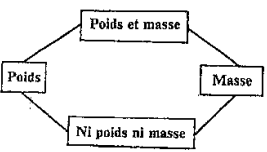
- La solution A (A)
- La solution B (B)
- Les deux (D)
- Aucune des deux (O).

Les Anglais appellent cette dernière solution « the null option », et nous pourrions l'appeler « le rejet ». Voici un exemple (D'Hainaut, 1973) :

Pour faire dévier un corps en mouvement en terrain plat, il faut lutter contre :

1. Son poids.
2. Sa masse.
3. Les deux.
4. Ni l'un ni l'autre.

Les quatre combinaisons du simplexe sont utilisées :



5. Les associations composées (Consigne QR=31).

Ces questions ont une formulation complexe (que nous déconseillons). En voici un exemple (Mazer, 1975).

Soit deux propositions X et Y. Répondre :

A. Si l'objet de la question est associé seulement avec X.
 B. Si l'objet de la question est associé seulement avec Y.
 C. Si l'objet de la question est associé à la fois avec X et Y.
 D. Si l'objet de la question n'est associé ni à X ni à Y.

Les questions concernent des signes cliniques que l'on observe dans les syndromes suivants :

X : syndrome cérébelleux.
 Y : syndrome tabétique.

Question 1 : Ataxie.
 Question 2 : Nystagmus.
 Question 3 : Dissociation des sensibilités.
 Question 4 : Signe de Babinski.
 Question 5 : Douleurs fulgurantes dans les membres inférieurs.

L'expression « association composée » vient du fait que chaque question doit être « associée » à une solution traitant elle-même de la nature des associations.

Fondamentalement, ces questions sont du type ABDO, avec mise en commun des solutions proposées pour plusieurs questions. Nous avons adapté la question de Mazer comme suit :

<p>Pour chacun des signes cliniques ci-dessous, dites s'il concerne :</p> <p>1. Le syndrome cérébelleux. 2. Le syndrome tabétique. 3. Les deux. 4. Ni l'un ni l'autre.</p> <p>Question 1 : Ataxie. Question 2 : Nystagmus. Question 3 : Dissociation des sensibilités. Question 4 : Signe de Babinski. Question 5 : Douleurs fulgurantes dans les membres inférieurs.</p>	
--	--

6. Les QCM A, B, Relation (Consigne QR=41).

Ces QCM portent sur deux éléments (A et B) et leur relation.

1° Assertion-raison

Voici une consigne et trois questions dues à Clarke (1975) :

Chacune des questions qui suivent consiste en une *assertion* et une *raison*.
 Déterminez si l'assertion est une proposition vraie ou fautive.
 si la raison est la proposition vraie ou fautive.
 si la raison est une explication correcte de l'assertion.

Ceci fait, écrivez :

Si vos réponses ci-dessus sont vrai, vrai, oui	1
vrai, vrai, non	2
vrai, faux	3
faux, vrai	4
faux, faux	5

	ASSERTION	RAISON	
Q1	L'hydrogène est utilisé pour remplir les ballons météorologiques	PARCE QUE	L'hydrogène est le gaz le moins dense de tous.
Q2	L'hydrogène pur ne brûle pas dans l'air	PARCE QUE	L'hydrogène pur est inerte.
Q3	L'hydrogène n'est habituellement ni préparé, ni obtenu par l'action du potassium sur l'eau.	PARCE QUE	le potassium réagit très peu avec l'eau.

Les trois propositions de base peuvent donner un simplexe de huit combinaisons, dont trois sans signification (d'où les cinq solutions restantes).

Au lieu de la QCM commune, il faudrait poser, pour chaque question, trois questions VRAI-FAUX, la troisième étant idiote dans le cas de réponses VF ou FV ou FF aux deux premières. Le remplacement de la formule « QCM AB relation » par « plusieurs VRAI-FAUX » devient donc moins évident.

Les réponses correctes aux trois questions de Clarke sont successivement 1 (niveau 3 du simplexe), 5 (niveau 0 du simplexe) et 3 (niveau 1 du simplexe).

Les sept solutions de la question X et les quatre solutions de la question Y peuvent être présentées dans deux simplexes opposés.

Dans l'exemple proposé, quelle que soit la réponse aux questions X et Y, on peut aussi demander à l'étudiant de répondre à la question de départ.

Ce n'est plus possible quand les données sont *insuffisantes* pour répondre. On ajoute alors aux solutions l'équivalent de « On ne peut répondre à la question : des données ou des informations manquent dans l'énoncé (l'amorce) ».

L'exemple ci-dessous est emprunté au Scholastic Aptitude Test, cité par Wood, 1977, p. 219.

Une quantité est indiquée dans la colonne A, une autre dans la colonne B. Comparez-les et indiquez :

1. Si la quantité figurant dans la colonne A est supérieure.
2. Si celle de la colonne B est supérieure.
3. Si les deux quantités sont égales.
4. Si la conclusion ne peut être tirée à partir de l'information fournie.

	Colonne A	Colonne B
Question 1	1	x
Question 2	$3 \times 352 \times 8$	$4 \times 352 \times 6$
Question 3	$\frac{x^2 - y}{y^2 - x}$	$\frac{x}{y} - \frac{y^2}{x^2}$

F. La Consigne « appariement » (Consigne QR=01).

Dans les questions à appariements, il s'agit de faire correspondre un élément d'une liste (les solutions proposées) à un élément de l'autre liste (les questions).

1) La correspondance biunivoque (forme déconseillée).

Consigne : « Choisissez, parmi les personnes indiquées à droite, le partenaire d'un personnage figurant à gauche.

Bloc de 4 questions :	Solutions :
Question 1 : Daphnis.	1. Cléopâtre.
Question 2 : Antoine.	2. Dalila.
Question 3 : Samson.	3. Juliette.
Question 4 : Roméo.	4. Chloé.

Les réponses correctes sont, 4, 1, 2, 3.

Dans cet ensemble de correspondances biunivoques, il suffit de connaître la réponse correcte à trois questions pour trouver automatiquement la dernière.

2) La forme asymétrique par excès (de solutions).

Consigne : « Parmi les solutions numérotées à droite, choisissez le partenaire de chaque personnage figurant à gauche. Attention ! Il y a plus de partenaires que nécessaire. »

Bloc de 4 questions :	Solutions :
Question 1 : Daphnis.	1. Cléopâtre.
Question 2 : Antoine.	2. Juliette.
Question 3 : Samson.	3. George Sand.
Question 4 : Roméo.	4. Dalila.
	5. Chloé.

Les réponses (R.C.) sont 5, 1, 4, 2. Dans cet exemple, la solution 3 est inutilisée.

Autre exemple (dû à J. Henry-Demol) :

Voici une liste de spécialités :	Pour chacune des spécialisés ci-dessous, indiquez celle de gauche qui a des effets semblables :
1. Acenterine.	Q. 109 : Cerebocine.
2. Bactrim.	Q. 110 : Eusaprim.
3. Cevitan.	Q. 111 : Neogynon.
4. Chymar.	Q. 112 : Novabritine.
5. Clamoxyl.	Q. 113 : Redoxon.
6. Neo-hombreol.	Q. 114 : Rhonal.
7. Percincamine.	
8. Stedril.	

Réponses correctes = 7, 2, 8, 5, 3, 1.

3) *La forme asymétrique par défaut (de solutions).*

Consigne : « Parmi les solutions numérotées à droite, choisissez la partenaire de chaque personnage figurant à gauche. Attention ! Pour un ou plusieurs des personnages de gauche, la partenaire correcte n'est pas proposée dans la colonne de droite. En pareil cas, il faut répondre par le code 9 »

<i>Bloc de 5 questions :</i>	<i>Solutions :</i>
<i>Question 1 : Daphnis.</i>	1. Cléopâtre.
<i>Question 2 : Antoine.</i>	2. Dalila.
<i>Question 3 : Samson.</i>	3. Juliette.
<i>Question 4 : Paris.</i>	4. Chloé.
<i>Question 5 : Roméo.</i>	

Les réponses correctes (R.C.) sont 4, 1, 2, 9, 3.

4) *La forme asymétrique par excès et par défaut (de solutions) (recommandée).*

Consigne : « Parmi les solutions numérotées à droite, choisissez la partenaire de chaque personnage (de gauche). Attention ! Pour certains personnages, la partenaire correcte ne figure pas parmi les solutions proposées. Il faut alors répondre 9. Il peut en outre arriver qu'une solution soit valable pour plusieurs questions ».

<i>Bloc de 4 questions :</i>	<i>Solutions :</i>
<i>Question 1 : Antoine.</i>	1. Dalila.
<i>Question 2 : Paris.</i>	2. George Sand.
<i>Question 3 : Chopin.</i>	3. Cléopâtre.
<i>Question 4 : Musset.</i>	4. Juliette.

Les réponses correctes (R.C.) sont 3, 9, 2, 2.

Pour les questions 3 et 4, la même solution (2. George Sand) est correcte ; les solutions 1 et 4 ne se rapportent à aucune question.

On voit que l'égalité entre le nombre de questions et le nombre de solutions est trompeuse. Le groupe de questions à appariements qui suit (dû à J. Henry-Demol) constitue un autre exemple de ce type de questionnement.

Voici une liste d'usages thérapeutiques :	Quel usage correspond à chacun des médicaments suivants ?
1. Anorexigène.	Q 190 : DIETHYPROPION
2. Anti-coagulant.	Q 191 : LEVO-DOPA
3. Anti-hémorragique.	Q 192 : MADRIBON
4. Anti-histaminique.	Q 193 : PARANAUSINE
5. Anti-infectueux.	Q 194 : PROMETHAZINE
6. Anti-parkinsonien.	Q 195 : SOUS-ACETAGE DE PLOMB
7. Révulsif.	Q 196 : TROMEXAN
8. Tranquillisant.	

Répondez par 9 si vous estimez que l'usage correct ne vous est pas proposé.

Pour remplacer les 7 QCM ci-dessus (à solutions communes) par plusieurs questions VRAI-FAUX, il faudrait en poser 8 par médicament, soit 56 au total !

Une telle débauche d'efforts est d'autant plus inutile que, dans ces exemples, les distrateurs sont plus utilisés dans le rôle de « masque de la réponse correcte » que comme des propositions intéressantes en elle-mêmes.

G. La consigne « association à terme exclu » (deux questions à consigne QR=11).

On peut présenter à l'étudiant des solutions combinées et, en plus, lui demander plusieurs réponses. L'exemple le plus connu de ce type de question est appelé « association à terme exclu ».

Exemple de Mazer, 1975 :

Voici, d'une part, une liste de trois énoncés et, d'autre part, cinq propositions numérotées. Quatre de ces cinq propositions sont communes à l'un (et à un seul) des énoncés. Pour répondre, il faut :

- cocher l'énoncé correspondant à ces quatre propositions,
- cocher la proposition rejetée (le terme exclu).

EXEMPLES

A. Accélération cardiaque.
B. Vasodilatation.
C. Ralentissement respiratoire.

- 1) Hémorragie.
- 2) Exercice musculaire.
- 3) Stimulation d'un nerf de Hering.
- 4) Section des nerfs pneumogastriques au cou.
- 5) Stimulation du ganglion stellaire.

Ce type de question constitue *la forme* la plus complexe des QCM : elle contraint à examiner chacune des propositions en se demandant si, pour le premier des trois « énoncés », elle constitue l'EXCEPTION. Si plusieurs exceptions se présentent, on recommence avec l'énoncé suivant.

En termes de questions VRAI-FAUX, il faudrait appliquer 3 questions VRAI-FAUX à chacune des cinq propositions (donc 15 questions) puis repérer l'énoncé ayant donné lieu à 4 VRAI, puis enfin repérer la (seule) proposition rejetée.

C'est en fait ce qui se passe en partie dans le cerveau de l'étudiant. En partie seulement, parce qu'un « pattern » de réponses s'impose à son esprit, avec des confirmations partielles mais suffisantes, avant d'avoir répondu à tous les VRAI-FAUX.

Le recours à cette consigne (association à terme exclu) est peu recommandable avant l'université car la compréhension de la consigne constitue à elle seule un problème. Nous coderons ceci comme *deux QCM différentes*.

H. La consigne « questions solidaires » (plusieurs QCM à consigne classique).

1) Les QCM en cascade

Pour obtenir le tarif prévu en cas de réponse correcte à une question de ce type, il faut fournir la réponse correcte à chacune des questions particulières, la réponse à l'une étant liée à la réponse de l'autre.

On utilise les expressions « QCM en cascade » ou « QCM à double ou à multiple face » ou « QCM emboîtées » pour désigner ces questions.

Question 1 : Quand c'est l'hiver (à Paris), la terre est :	RC - 1
1. plus proche	
2. plus éloignée	
3. à égale distance	
du soleil que lorsque c'est l'été (à Paris)...	
Question 2 : ... parce que la trajectoire de la terre autour du soleil est :	RC - 3
1. parfaitement circulaire	
2. très elliptique (forme très allongée)	
3. légèrement elliptique (forme très proche du cercle)	
Question 3 : ...et que le soleil se trouve	RC - 2
1. au centre	
2. au foyer	
de cette trajectoire ...	
Question 4 : ... si bien que la distance terre-soleil	RC - 3
1. est le facteur <i>unique</i>	
2. est un facteur important	
3. n'est pas un facteur	
qui explique la différence de température entre l'hiver et l'été.	

Ces quatre questions relatives à un thème unique, sont typographiquement liées entre elles par les points de suspension et isolées des autres par un cadre en traits pleins ou par le mot « question ».

Les questions 2 et 3 portent sur l'explication de la réponse en 1, tandis que la question 4 porte sur un des développements de celle-ci.

Il est fréquent que, dans des questions solidaires, certaines portent sur la *cause*, la *conséquence*, le *but*, les *circonstances*, les *comparaisons*, les *restrictions*, etc, que l'on peut associer à la réponse à une autre question.

On peut rendre les différentes questions solidaires par leur notation et n'attribuer de points que si les quatre questions sont *toutes* réussies.

On constate que les chances de trouver par hasard la série de réponses correctes sont dans notre exemple réduites à $1/3 \times 1/3 \times 1/2 \times 1/3 = 1/54$. Certains partent de ce raisonnement pour recourir, chaque fois qu'ils le peuvent, à une telle notation solidaire. Mais cette pratique n'offre pas que des avantages : l'analyse expérimentale des questions s'en trouve notablement compliquée.

*
* *

2) QCM partiellement solidaires.

Dans un ensemble de questions, certaines peuvent être *sélectivement* solidaires avec d'autres, comme dans l'exemple qui suit (De Landsheere, 1980, p. 103 sq), intitulé *Analyse d'observation*, et qui met le candidat dans une situation comparable à l'expérience réelle.

« *Exposé de la maladie* : Le malade est un homme de 21 ans qui se plaint de malaises, d'une toux et de fièvre.

La maladie a débuté dix jours avant l'admission par un malaise et une toux sans expectoration, suivis dans les 24 heures d'une température variant de 37,8 à 38,3, qui a persisté jusqu'au moment de l'admission.

Le quatrième jour de la maladie, la toux s'accroît, produisant de petites quantités d'expectoration blanche et visqueuse.

Trois jours avant l'admission, des accès paroxystiques de toux commencent, parfois suivis de vomissements. Des sensations de frissons furent notées, mais non pas de véritables frissons avec tremblements. Une douleur parasternale antérieure à la toux existe depuis le cinquième jour de la maladie.

A l'examen physique, la température est à 38,3, le pouls à 110, le rythme respiratoire 32, la tension maximum 10 1/2, minima 8. Le malade est bien développé, sans maigreur, sa maladie semble aiguë, il est dyspnéique, mais non cyanosé.

L'examen physique de la cage thoracique montre des vibrations vocales à la palpation et à l'auscultation. Le murmure vésiculaire est normal. Dans l'aisselle gauche, on entend quelques râles fins et la qualité bronchique du son est augmentée, bien que d'intensité normale.

La formule sanguine est la suivante : globules blancs : 3 400 (polynucléaires 30 %, lymphocytes 62 %, monocytes 5 %, éosinophiles 3 %).

La radio du thorax révèle une augmentation de la densité de la région périciliaire avec des aires mal définies de densité inégale, nuageuse aux deux bases et dans un champ pulmonaire supérieur gauche.

Questions :

1. Quel est le diagnostic le plus probable ?
 - a) Tuberculose.
 - b) Pneumonie à pneumocoques.
 - c) Pneumonie (primaire atypique) à virus.
 - d) Coccidiomycose.
 - e) Broncho-pneumonie.
2. Quel est le signe physique qui s'y ajoute probablement ?
 - a) Spénomégalie.
 - b) Signe de souffrance méningée.
 - c) Bruit de frottements pleuraux.
 - d) Changements fréquents dans la distribution des symptômes thoraciques.
 - e) Signe de condensation lobaire gauche.
3. Lequel des examens de laboratoire suivants va de pair avec le diagnostic ?
 - a) L'élévation et l'augmentation des agglutines froides.
 - b) Hémoculture positive.
 - c) Leucocytose marquée au début de la convalescence.
 - d) Examen des expectorations.
 - e) Cuti-réaction positive.
4. Quelle thérapeutique devra être employée ?
 - a) Repos au lit et streptomycine.
 - b) Repos au lit et pénicilline.
 - c) Streptomycine et acide para-aminosalicylique.
 - d) Repos au lit et auréomycine.
 - e) Psychothérapie et rééducation physique.
5. Quelle est l'issue probable de cette maladie, sans traitement ?
 - a) La fièvre va disparaître spontanément par une crise terminale.
 - b) La convalescence va être progressive avec une rechute prévisible.
 - c) Un emphysème résiduel va se développer.
 - d) Une fibrose résiduelle va apparaître après guérison.
 - e) Une caverne pulmonaire peut apparaître. »

Dans cette série de questions à (longue) introduction commune, la première a une importance cruciale : l'étudiant qui se trompe à la question 1

se trompera vraisemblablement dans les quatre autres. La réponse à la question 1 conditionne la réponse aux suivantes. La réponse à la question 1 était correcte. Donc, en cas d'erreur à la question 1, il est impossible d'obtenir un seul point.

Les professeurs de médecine qui recourent à ce système de notation mettent en avant l'importance d'un diagnostic correct pour justifier une telle sévérité.

3. Les codes concernant la liaison entre questions (L)

Le code $L = 0$ représente la situation la plus courante où chaque question est corrigée indépendamment des autres.

Les codes $L = 1$, $L = 2$ et $L = 3$ représentent la liaison de plusieurs réponses au niveau de la cotation : plusieurs réponses doivent être *simultanément* correctes pour qu'un tarif donné soit appliqué.

a) Liaison entre questions « double face »

Question N

Q = 1 une seule solution correcte

R = 1 fournissez une seule réponse ou omettez.

L = 1 question liée aux suivantes

(1 indique qu'il s'agit de la première des questions liées entre elles)

Question O

Q = 1 une seule solution correcte

R = 1 fournissez une seule réponse ou omettez.

L = 3 question liée aux précédentes

(3 indique qu'il s'agit de la dernière des questions liées).

Question N	H Cl est le symbole
	1. du chlorure d'hydrogène,
	2. de l'hydruure de chlore,
	3. du chlorate d'hydrogène,
	4. du perchlorate d'hydrogène.
Question O	et, en solution aqueuse, il s'ionise en donnant
	1. des ions H^+ (ou H_3O^+) et Cl^-
	2. des ions H^+ (ou H_3O^+) et Cl^-
	3. des ions H^+ (ou H_3O^+) et Cl^-
	4. des ions H^+ (ou H_3O^+) et Cl^-

Le tarif prévu pour une réponse correcte (TC) ne sera attribué que si la réponse 1 est donnée à la question N et la réponse 4 à la question O. La

note (omission) ne sera accordée que s'il n'est répondu à aucune des deux questions. Tout autre ensemble de réponses est considéré comme incorrect.

b) Liaisons dans un VRAI-FAUX généralisé

On peut aussi exiger (code $L > 0$) que les k solutions d'une QCM aient *toutes* (code $R = 4$) fait l'objet d'une réponse correcte (VRAI-FAUX). On dira alors de ces k solutions qu'elles sont des *sous-questions solidaires* au niveau de la notation. Elles reçoivent un code QRL chacune.

Voici un exemple (sous-questions P1 à P9) dû à Debot :

Variante « code L = 0 »			Variante « code L = 0 »		
Q	R	L	Q	R	L
1	4	1	1	1	0
1	4	2	1	1	0
1	4	2	1	1	0
1	4	2	1	1	0
1	4	2	1	1	0
1	4	2	1	1	0
1	4	2	1	1	0
1	4	3	1	1	0

Indiquez, pour chacune des couleurs suivantes, si elle appartient (VRAI ou FAUX) à la catégorie des couleurs spectrales primaires.	
1. l'infra-rouge	(V)
2. le rouge	(V)
3. l'orange	(V)
4. le jaune	(V)
5. le vert	(V)
6. le bleu	(V)
7. l'indigo	
8. le violet	
9. l'ultra-violet	

Avec la variante $L > 0$, le tarif attribué à une réponse correcte (TC) ne sera accordée que pour la série de réponses FVFFVFFFF, tandis qu'avec la variante $L = 0$, chaque sous-question est considérée comme une question indépendante et corrigée comme telle.

Dans le cas du jugement multiple (code $R = 4$), comme pour les questions P1 à P9 ci-devant, chaque solution doit avoir *son* poids. Dans la variante « code $L > 0$ », les poids des questions liées sont additionnés.

c) Liaison entre plusieurs QCM

Une formule originale de combinaison de plusieurs QCM a été développée par Halleux (1969) qui focalise l'attention de l'élève sur la détection d'erreurs dans les diverses questions portant sur un énoncé commun. Il s'agit de repérer à *quel endroit précis se situent les erreurs*.

Chaque sous-question décrit, en plusieurs lignes, les étapes du calcul. Dans certaines sous-questions, l'auteur *a introduit une erreur* et l'élève doit désigner dans quelle ligne. N'oublions pas qu'une erreur se répercute dans les lignes qui suivent celle où elle est apparue. Une seconde erreur (inverse) serait nécessaire pour faire disparaître la première. C'est pourquoi, dans la consigne, on demande aux étudiants de désigner, non pas la ligne contenant une erreur, mais *la ligne où l'erreur est apparue pour la première fois*.

Questions dues à Halleux :

Effectuer $(x + 5)^2 - (x - 5)^2$

Consigne : Notez, pour chaque question Q1 à Q5, dans quelle ligne est apparue l'erreur (s'il y en a une) pour la première fois. Si le raisonnement est sans erreur, répondez par k+1 (aucune).

	Variante « Code L ≥ 0 »			Variante « Code L = 0 »		
	Q	R	L	Q	R	L
Q1	1	1	1	1	1	0
Q2	1	1	2	1	1	0
Q3	1	1	2	1	1	0
Q4	1	1	2	1	1	0
Q5	1	1	3	1	1	0

$(x + 5)^2 - (x - 5)^2$
 Q1 1. = $x^2 + 10x + 25 - x^2 - 10x + 25$
 2. = 50
 Q2 1. = $x^2 + 25 - x^2 + 25$
 2. = 50
 Q3 1. = $x^2 + 10x + 25 - x^2 + 10x - 25$
 2. = 20x
 Q4 1. = $x^2 + 5^2 - x^2 - 5^2 + 10x + 10x$
 2. = 20x
 Q5 1. = $(x + 5 - x + 5)(x + 5 + x - 5)$
 2. = 10 · 2x
 3. = 20x

I. Discussion

Toutes les formes de QCM envisagées jusqu'ici permettent de répondre par *un* choix, car *une seule solution est correcte*. Les premières formes de QCM peuvent être remplacées par des séries de questions VRAI-FAUX, procédé plus analytique mais moins économique. Ce remplacement devient malaisé, voire excessivement coûteux, pour des formes sophistiquées : QCM A+B+Relation à appariement, association à terme exclu. A cette extrémité du continuum, les distracteurs apparaissent plus comme des « masques » à la réponse correcte que comme des propositions sur lesquelles il importe de recueillir des informations.

C'est en dernier ressort ce qui guidera notre choix : sur quelles propositions voulons-nous absolument mesurer la connaissance de l'étudiant, et quel prix (en temps) sommes-nous prêts à payer (et à lui faire payer) pour cela ?

CHAPITRE 3

La rédaction des QCM

- A. Styles de rédaction des QCM simples
 1. Le style direct.
 2. Le style indirect.

Règles de rédaction concernant...

- B. ... L'adéquation aux objectifs
 1. Respecter l'objectif.
 2. Coller à l'objectif.
 3. Ne pas perturber des apprentissages.

- C. ... La valeur diagnostique de la réponse
 4. Révéler le processus mental.
 5. Indiquer l'erreur commise.
 6. Préciser les lacunes.

- D. ... La forme
 7. Respecter la consigne.
 8. Respecter la syntaxe.
 9. Pas de terme vague.
 10. Peu de négation.
 11. Séparer informations et questions.
 12. Regrouper dans l'amorce.

- E. ... Les solutions proposées
 13. Indépendance syntaxique.
 14. Indépendance sémantique.
 15. Mêmes mots communs avec l'amorce.
 16. Même vraisemblance.
 17. Même longueur.
 18. Même complexité.
 19. Même degré de généralité.
 20. Même degré de technicité.

F. L'habileté à répondre à des QCM

1. Définition.
2. Les méthodes d'étude de cette habileté.

G. Résultats d'études expérimentales

1. Transgression de la règle 8.
2. Transgression de la règle 14.
3. Transgression de la règle 15.
4. Transgression de la règle 16.
5. Transgression de la règle 17.
6. Transgression de la règle 19.
7. Transgression de la règle 20.
8. L'ordre des solutions correctes dans un test.
9. Observations générales.

Introduction

Ecrire une QCM est plus délicat qu'il n'y paraît à première vue. Pour rédiger ou juger des QCM, il est utile de disposer de critères précis.

Divers auteurs (Travers, 1955; Gronlund, 1965; Hedges, 1971; etc.) fournissent des listes de conseils pour la rédaction de QCM. La plupart combinent deux types de recommandations: ce qu'il convient d'éviter et ce qu'il convient de faire.

Nous distinguerons quatre catégories de règles:

Trois règles d'adéquation aux objectifs.

Trois règles sur la valeur diagnostique.

Six règles générales de rédaction des QCM et d'autres types de questions.

Huit règles de rédaction des solutions proposées.

La violation de l'une des six premières règles met en cause la pertinence des mesures obtenues et la possibilité de les interpréter de façon correcte.

Ne pas respecter les six règles suivantes rend la question plus difficile, tandis qu'une infraction aux huit dernières règles la rend plus facile. Ce type de faille dans une QCM met en effet sur la voie de la réponse correcte. Certains étudiants sont particulièrement habiles à détecter ces indices. Cette compétence, appelée *test-wiseness* en anglais, sera étudiée en fin de chapitre.

A. Styles de rédaction des QCM simples

Le style de rédaction peut influencer la difficulté de la question, voire en changer la portée ou le contenu.

1. Le style direct (partant de l'amorce)

Une QCM directe (en anglais forward item) est une question à laquelle on pourrait répondre sans avoir à lire les solutions proposées, bref une question ouverte à laquelle on aurait ajouté des solutions proposées comme dans les deux exemples ci-dessous.

<p>Quelle est la capitale du Mali ?</p> <ol style="list-style-type: none"> 1. Tombouctou 2. Ouagadougou. 3. Bamako.
--

RC = 3

<p>Quelle est (ou quels sont) le(s) fleuve(s) qui coule(nt) en Côte d'Ivoire ?</p> <ol style="list-style-type: none"> 1. Le Bandama. 2. Le Comoué. 3. Leassandra. 4. La Casamance.
--

RC = 1, 2 et 3.

Dans la majorité des cas, la réponse à la QCM est plus facile que la réponse à la question ouverte correspondante, et ce pour trois raisons :

1. La présence de la ou des solutions correctes parmi les solutions transforme la performance d'évocation (appelée par une question ouverte) en une performance de reconnaissance, plus facile.
2. La réponse correcte étant déjà formulée, l'étudiant est à l'abri de certaines erreurs (comme les fautes d'orthographe, par exemple).
3. L'auteur de la question a déjà choisi, la ou les solutions correctes qui lui conviennent.

Les amorces des QCM présentées de manière interrogative auraient pu l'être sous forme de phrase lacunaire.

Exemple : La capitale du Mali est...

La valeur du rapport vaut...

L'amorce aurait pu se décomposer en une phrase de présentation d'abord et en une phrase interrogative ensuite.

Exemple : Voici des villes d'Afrique. Laquelle est...

Voici des coefficients. Lequel représente...

De telles variations superficielles combattent la monotonie, mais n'ont pas d'influence sensible sur la facilité de la question, la dernière forme (Voici...) convient mieux lorsque l'amorce est longue : il est préférable de la fractionner en plusieurs phrases distinctes pour en améliorer la lisibilité (voir Henry, 1977).

2. Le style indirect, partant des solutions

Ces QCM ont parfois été appelées « questions à rebours » (backward items) parce que, pour y répondre, l'étudiant est obligé d'examiner les solutions proposées. Il ne s'agit donc plus d'une question ouverte à laquelle on aurait ajouté des solutions. Néanmoins, comme les questions directes, leur structure profonde relève du simplexe décrit ci-avant. Dans la forme la plus répandue de ce type de QCM, on demande de choisir, parmi les solutions proposées, celle(s) qui ne répond(ent) pas aux conditions décrites dans l'amorce ; ce sont les QCM indirectes du type « SAUF » ou « EXCEPTE ».

On rencontrera, en outre, parmi les QCM indirectes, des QCM avec amorce complète, des QCM sans amorce et des QCM avec amorce incomplète (ou pseudo-amorce).

a) QCM indirectes du type « sauf » ou « excepté »

Voici un exemple de ce type de question :

<p>En termes botanique, tous les fruits suivants sont des pommes, EXCEPTE :</p> <ol style="list-style-type: none"> 1. Les poires. 2. Les pommes. 3. Les grenades. 4. Les fruits de l'aubépine. 	<p>RC = 3</p>
--	---------------

Il serait impossible de répondre à la question si les solutions proposées étaient absentes.

En principe, il ne devrait pas être plus difficile de répondre à une telle QCM portant sur une restriction, une exception, qu'à une QCM classique formulée positivement. Cependant, l'amorce négative rend la lecture légèrement plus ardue et accroît le risque d'erreur par distraction.

L'augmentation de la difficulté due à l'emploi de tournures négatives a été mise en évidence expérimentalement par Terranova (1969), Dudycha et Carpenter (1973).

Ce type de question suppose que l'on présente les combinaisons de niveau 1 du simplexe des sous-questions.

Il peut arriver qu'une QCM du type « Sauf » soit une QCM directe (que l'on puisse y répondre avant de lire les solutions), comme dans l'exemple suivant :

RC = 3

Le noyau d'un atome est formé de protons et de neutrons, pour tous les corps, SAUF pour :

1. le lithium
2. le bore
3. l'hydrogène
4. l'oxygène.

b) QCM indirectes avec amorce complète

Exemple :

RC = 1

Laquelle des villes ci-dessous est située au Mali ?

1. Tombouctou.
2. Niamey.
3. Lomé.
4. Konakri.

c) QCM indirectes sans amorce

Exemple :

RC = 2

Parmi les affirmations suivantes, laquelle est correcte ?

1. Lors d'une inspiration, nos poumons absorbent 50 % de l'oxygène de l'air inhalé.
2. L'azote inspiré est égal, en volume, à l'azote expiré.
3. La respiration consiste à échanger l'azote et l'oxygène.

L'amorce ne présente pas de problème précis. Cette formulation cache le plus souvent le regroupement abusif de plusieurs questions VRAI-FAUX, qu'il eût mieux valu laisser distinctes.

d) QCM indirecte avec pseudo-amorce

De telles amorces ne font que situer le contexte, mais ne posent pas la question.

Exemple (Clarke 1975, p. 4) :

RC = 2

Il est exact de dire que quand l'hydrogène brûle dans l'air :

1. il se produit une oxydation de l'hydrogène.
2. la réaction est exothermique.
3. l'hydrogène alimente la combustion.
4. il se produit une réduction de l'oxygène.

B. Règles de rédaction concernant l'adéquation aux objectifs

REGLE 1

N'utiliser la QCM que si elle est le type de question le plus approprié à mesurer ce que l'on vise

Cette règle est fondamentale. Nous y consacrerons beaucoup d'attention. Dans les exemples ci-dessous, une *autre* modalité de questionnement doit être préférée aux QCM.

a) La question ouverte par intervalle

Au chapitre I (section C2), nous avons déjà indiqué quelles situations (dates,...) appellent un tel mode de questionnement.

b) La question VRAI-FAUX

La question ci-dessous est un exemple de QCM que l'on remplacerait heureusement par quatre questions VRAI-FAUX. Remarquons que l'amorce de cette question ne présente pas le problème.

1b

La Chine

1. est communiste.
2. a une superficie de 9 580 000 km².
3. a 800 millions d'habitants au moins.
4. a pour capitale PEKIN.

La présence de la solution 3 est de nature à alerter les élèves les moins perspicaces, et donc à enlever à la question tout son pouvoir révélateur. Nous verrons, au chapitre 4, section C, une possibilité de poser de telles questions au moyens de QCM, sans mettre « la puce à l'oreille » de l'étudiant. Cette possibilité recourt aux solutions dites générales. En dehors de cette procédure, la question ouverte paraît s'imposer.

4^e Mesure de la capacité d'invention

L'exemple suivant porte sur la créativité, la divergence.

II

Inventez un nom de poudre à lessiver en insistant sur l'une de ses qualités : mousse abondante, propreté du linge, peu coûteuse, enlève toutes les tâches... Choisissez le nom qui vous plaît le mieux.

1. KIMOUS
2. ANITACH
3. SANTACH
4. IMPEC
5. MOUSPACHER
6. ECONOMOUS
7. MOUSTACH

La présentation de solutions déjà rédigées ne permet pas de mesurer les qualités visées dans cette question, à savoir



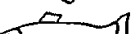
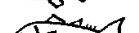

- la fluidité verbale (capacité à fournir beaucoup de termes en un temps donné) ;
- la flexibilité (capacité à changer de registre, à varier les domaines considérés) ;
- l'élaboration (capacité à structurer, à construire sa réponse, à l'assortir de détails pertinents).

Les trois aspects ci-dessus sont habituellement codés par les spécialistes de l'évaluation de la créativité à partir de réponses rédigées. Les tests en la matière (ceux de Torrance par exemple) sont constitués de questions ouvertes.

REGLE 2

La QCM doit correspondre à l'objectif visé, au comportement à évaluer

Une même matière peut faire l'objet de plusieurs questions qui, bien que proches l'une de l'autre, sollicitent des comportements différents parce qu'elles visent des objectifs différents (exemple ci-dessous).

<p>2a</p> <div style="border: 1px solid black; padding: 5px;"> <p>Où est la tanche ?</p> <ol style="list-style-type: none"> 1.  2.  3.  4.  </div> <p>RC - 2</p>	<p>2b</p> <div style="border: 1px solid black; padding: 5px;"> <p>Ce poisson est</p>  <ol style="list-style-type: none"> 1. un brochet 2. une truite 3. une carpe 4. une tanche </div> <p>RC - 4</p>
--	---

La question 2a correspond à la situation du *poissonnier* qui doit reconnaître, parmi les diverses espèces présentes sur son étal, le poisson demandé par le client.

La question 2b représente la situation du *pêcheur* à la ligne qui, même dans le meilleur des cas, n'a qu'un seul poisson au bout de sa ligne et doit lui donner un nom.

Selon la capacité que l'on veut tester, on posera la question 2a ou la question 2b.

REGLE 3

La QCM ne doit pas perturber des apprentissages

Dans les domaines où la mémoire perceptive importe, il faut éviter les distracteurs pouvant se fixer chez l'étudiant. Ainsi, la question 3a ne peut pas être présentée sous forme de QCM.

3a.

Choisissez l'orthographe correcte :	
1. Orquidée	5. Orchidée.
2. Horquidée.	6. Horchidée.
3. Orkidée	7. Horchydée.
4. Horkidée.	

Cette incompatibilité paraît évidente. Pourtant, aux Etats-Unis, de nombreux tests d'orthographe d'usage ont été construits selon ce principe. Par exemple, dans le test d'orthographe de RICH et ENGELSON (édité à Emporia au Kansas), les étudiants doivent indiquer dans une liste de mots, lesquels sont orthographiés incorrectement. Ils sont ainsi exposés aux graphies (incorrectes) suivantes :

« évuidance, providance, bazzar, acepting, fragrence, indefinetely ».

Souvent, le problème se pose en termes plus nuancés. Ainsi, la question 3b ne doit pas être posée à des enfants en période d'apprentissage. Il n'en irait pas de même pour sélectionner des correcteurs de journaux.

Il est souvent possible de formuler la question de manière à supprimer cet inconvénient. Ainsi, la question ci-dessous (3b).

3b

Quelle phrase ne contient pas de faute d'orthographe ?
1. Je vous ai fait patienté.
2. Vous êtes arrivéz.
3. J'ai jouer.
4. Vous venez.

peut être facilement transformée comme suit :

Complétez les mots mutilés par : 1 = é ; 2 = er ; 3 = ez.	
3c.	Je vous ai fait patient —
3d.	Vous êtes arriv —
3e.	J'ai jou —
3f.	Vous ven —

Des précautions s'imposent, non seulement dans le domaine de la perception visuelle, mais aussi pour les perceptions auditive, tactile, musculaire, etc.

Ainsi, prononcer de trois façons différentes (dont deux incorrectes) le verbe néerlandais *regeren* (gouverner) créerait une regrettable confusion chez un étudiant débutant. N'arrive-t-il pas, même à des professeurs de douter de l'orthographe d'un mot à force de le voir mal écrit ?

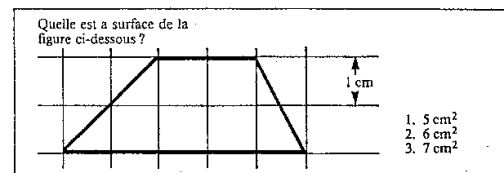
C. Règles de rédaction concernant la valeur diagnostique de la réponse

REGLE 4

La QCM doit renseigner le professeur sur le processus mental utilisé par l'étudiant

a) Examinons la question 4a

4a



Il est possible d'y répondre de deux manières, au moins :

- 1) par la méthode enseignée en classe :
 - mesurer les bases de la hauteur du trapèze.
 - mesurer la hauteur.
 - calculer la surface par la formule $S = \frac{H}{2} (B+b)$
- 2) en comptant les carrés d'un cm².

b) La question 4b constitue un autre exemple.

4b	Quelle est l'aire d'un cercle de 6 cm de diamètre ?	
	1. 9,42 cm ²	3. 28,26 cm ²
	2. 18,84 cm ²	4. 56,52 cm ²

Il est possible d'y répondre

- 1) par la méthode à laquelle le professeur pense : appliquer la formule de l'aire du cercle (πR^2)
- 2) dessiner le carré qui inscrit exactement ce cercle : constater que ce carré a 6 cm de côté, donc une aire de 36 cm². Choisir la solution 3.

La deuxième performance est plus astucieuse que la première, mais permet de l'éviter, ce qui serait impossible si les solutions proposées étaient plus « proches » les unes des autres.

c) La question 4c est un exemple classique du même problème.

4c	Que vaut $\sqrt{234,3961}$?	
	1. 15,31	3. 15,43
	2. 15,30	4. 15,29

Il est possible d'y répondre de deux façons

- 1) en extrayant la racine carrée (comme prévu par l'enseignant) ;
- 2) en éliminant les solutions 2 et 3 (car des nombres se terminant par 0 ou par 3 ne peuvent donner un carré se terminant par 1), puis en élevant la solution 1 au carré (pour voir).

d) La question 4d fournit un dernier exemple du même genre.

4d

Un homme et un jeune enfant poussent, chacun de leur côté, la barre rigide d'une balançoire. Cette balançoire reste immobile (voir dessin).

Quelle relation est correcte ?

1. $F_1 = F_2$
2. $F_1 \ll F_2$
3. $F_1 \gg F_2$



Il est possible d'y répondre de deux façons :

- 1) en appliquant la loi $F_1/F_2 = D_2/D_1$
- 2) en se disant qu'un adulte est plus fort qu'un enfant, donc que $F_1 > F_2$

Il aurait suffi au constructeur de la question de mettre en présence deux enfants ou deux adultes pour faire de cette question un problème portant vraiment sur le moment d'une force.

*
* *

Dans le même ordre d'idées, WOOD (1977) dénonce « les questions de laboratoire auxquelles on peut répondre sans avoir fait l'expérience » et « les questions de langues modernes auxquelles on peut répondre sans connaître la langue ».

Pyrzszak (1972 et 1974) critique « les questions de compréhension de texte auxquelles on peut répondre sans avoir lu le texte » (par exemple, parce que ce sont des questions de connaissance). Cet auteur suggère d'ailleurs une présentation des questions en deux temps : avant et après la lecture du texte. Une question qui serait réussie de la même façon dans les deux cas devrait être éliminée.

REGLE 5

Les distracteurs doivent indiquer le type d'erreur commise ou le cheminement incorrect suivi par l'élève

Powell (1968) regrette que l'on considère souvent le distracteur comme un « masque » de la réponse et que tous les distracteurs soient ramenés à la catégorie « réponse incorrecte ». La notation ne dépend, la plupart du temps, que du nombre de réponses correctes à l'ensemble du test et non des solutions incorrectes choisies. Cependant, dans un test bien construit, certaines solutions incorrectes peuvent atteindre un haut niveau de discrimination. Powell considère que le choix des distracteurs *ne se fait pas au hasard*, mais porte une signification.

Il a démontré par l'absurde, l'existence de ce phénomène. Ses hypothèses (absurdes) sont :

Hypothèse 1 : Quand on demande aux étudiants d'indiquer les raisons du choix d'une solution à une question donnée, les explications varient beaucoup d'un étudiant à l'autre.

Hypothèse 2: Bien que l'on observe un lien (factoriel) entre certains distracteurs aux différentes questions, on ne trouve pas de cohérence entre les raisons *du choix* de ces distracteurs avancés par les étudiants (à l'intérieur d'un même facteur).

Powell avance encore quatre autres hypothèses absurdes de ce genre que les données viennent contredire.

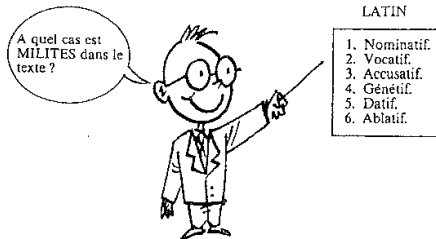
Dans les questions qui suivent, le choix minutieux des distracteurs par les constructeurs du test a pour effet, non seulement de rendre la solution correcte peu repérable, mais aussi de faire percevoir le processus mental utilisé par l'étudiant.

a) Les solutions naturelles

Pour certaines matières, les réponses possibles sont bien connues à l'avance. Ainsi, tout étudiant sait qu'il existe quatre groupes sanguins principaux (A, B, O, AB), qu'en allemand, le genre d'un nom commun ne peut être que masculin, féminin ou neutre, qu'en histoire deux événements ne peuvent être que dans une relation chronologique d'antériorité, de postériorité ou de contemporanéité.

Quand la question porte sur ces notions, peu importe qu'elle soit ouverte (réponse rédigée) ou à choix multiple : l'étudiant évoque de toute façon les diverses solutions possibles. Ces solutions sont invariables et « imposées » par la nature du contenu. C'est pourquoi nous les appelons QCM « naturelles ».

Ces questions ont d'ailleurs donné lieu à une pratique astucieuse, qui consiste à afficher, *avec un code*, ces solutions bien connues. On peut alors « improviser » des QCM *orales* élémentaires (comme dans l'exemple ci-dessous) :



C'est un des très rares cas où l'on peut improviser une QCM et la présenter (partiellement) de façon orale.

b) Les solutions inspirées des erreurs habituelles

L'idéal consiste à ne pas « inventer » des distracteurs, mais à utiliser ceux que les étudiants fournissent eux-mêmes quand la question est ouverte.

Exemple 1

La question 5a ci-dessous a été créée par Vandeveldt et Vanderelst (1975) pour illustrer le niveau « compréhension » de la taxonomie de Bloom et al. (1976). Cette QCM est destinée à des étudiants qui ne connaissent pas le grec ancien.

5a

Un spécialiste en oto-rhino-laryngologie indique généralement sur sa plaque « nez-gorge-oreilles ».
Le rhinocéros a une corne à l'avant du museau.
Celui qui souffre de rhinite se mouche fréquemment.

Le mot grec RHIS, RHINOS signifie :

1. Maladie.	4. Corne.
2. Visage.	5. Front.
3. Nez.	6. Gorge.

RC = 3

Les solutions *maladie, visage, corne, front* sont bien celles qui viennent à l'esprit du lecteur lorsque la question n'est pas à choix multiple. C'est pourquoi les constructeurs ont, *a priori*, utilisé ces distracteurs.

La présentation *sous forme ouverte* de cette question à des étudiants de quatorze ans ne connaissant pas le grec a révélé un grand nombre de réponses « gorge » non prévues au départ. A l'analyse, on s'est rendu compte que cette réponse découle du *raisonnement* suivant :

Rhino est le deuxième mot de l'expression « oto-rhino-laryngologie » ; il correspond donc au deuxième mot de « nez-gorge-oreilles ».

Exemple 2

Les solutions proposées dans la question 5b ont été reprises d'une liste d'erreurs publiées par Bonboir (1960) dans le compte rendu d'une vaste recherche pédagogique sur l'arithmétique au niveau de la sixième primaire.

5b

Ecrivez le pourcentage suivant sous forme de fraction :

6,75 % = $\frac{\quad}{\quad}$

1. $\frac{675}{100}$	3. $\frac{6,75}{100}$	5. $\frac{0,0675}{100}$
2. $\frac{67,5}{100}$	4. $\frac{0,675}{100}$	

RC = 3

- Dans l'étude de Bonboir, on peut trouver, pour chaque exercice :
- le taux d'erreurs ;
 - la liste des différentes erreurs et leurs proportions ;
 - la démarche suivie (connue par réflexion parlée directement enregistrée) par des élèves ayant réussi et par des élèves ayant échoué.

Ce genre de document est précieux pour le constructeur d'épreuves pédagogiques, car il suggère des distracteurs efficaces et il met sur la piste des remédiations¹.

Exemple 3

Dans l'exemple suivant, les solutions proposées sont, elles aussi, inspirées des types d'erreurs commises fréquemment par les élèves. Cette question (5c) ci-dessous est reprise au Test CL4 de Consultation du Dictionnaire de Roller (Genève - Service de la Recherche Pédagogique).

Les étudiants dispoisaient des pages 737 à 740 du petit dictionnaire Larousse illustré. L'extrait de la page 740 reproduit ci-dessous est suffisant pour permettre de fournir la réponse correcte (solution 4).

5c **PARACHUTISTE** n. Sportif ou militaire entraîné au saut en parachute. « ad. Trouver parachutistes, unités de parachutistes entraînés à combattre en petits groupes (commandos) ou dans le cadre d'une grande unité aéroportée. »

PARACLET [paraklet] n. m. (gr. *paraklētōs*, avocat). Titul. Non donné au Saint-Esprit.

PARADE n. f. (de *parer*). Rassemblement d'unités militaires pour les passer en revue ou les faire évoluer. | Action, manœuvre de parer un coup; défense, riposte. | Scène burlesque jouée à la porte d'un théâtre forain pour attirer le monde. | Équid. Arrêt brusque d'un cheval au manège. | Étiol. Comportement d'insolence envers les rivaux, ou comportement démonstratif envers le partenaire sexuel. | Littér. Symbole représenté dans les salons au XVII^e, et imitant les types et le langage populaires. « De parade, pour l'ostentation, plus que pour l'utilité. | Faire parade de, faire étalage d'une qualité afin de se faire valoir. | Être de parade. Il sur lequel on expose après leur mort les hauts personnages. »

PARADER v. l. se donner un air avantageux pour attirer l'attention. | Mlt. Manœuvrer, évoluer.

La paracousie est une sorte de surdité. Ce mot se trouve-t-il dans le dictionnaire qui a fourni les quatre pages que vous consultez en ce moment ?

1. Non, parce que le mot « paracousie » devrait se trouver à la page 737 et que je ne l'y trouve pas.
2. Oui, parce qu'un dictionnaire contient tous les mots de la langue française.
3. Non, parce que le mot « paracousie » devrait se trouver à la page 739 et que je ne l'y trouve pas.
4. Non, parce que le mot « paracousie » devrait se trouver à la page 740 et que je ne l'y trouve pas.
5. Je ne peux pas le savoir parce que je n'ai pas le dictionnaire entier sous la main.

¹ Comme, par exemple le Programme Auto-Correctif à Embranchements sous forme de Livre Brouillé à Réponses Ouvertes (Pacelbro). Voir Leclercq (1968, 1970).

REGLE 6

On doit préciser sur quelle partie de l'énoncé porte la question

Sans une telle précision, le diagnostic risque d'être complètement erroné. Voici quelques exemples de questions imprécises.

6 a

Au VIII^e siècle, Charles Martel a battu les Arabes à Poitiers, les chassant ainsi d'Europe.

RC - 2

1. Vrai.
2. Faux.

Seule la dernière partie de la proposition est fautive (car les Arabes sont encore restés plusieurs siècles en Espagne après cette bataille).

6 b

Les visons ont une fourrure épaisse parce qu'on en fait des manteaux.

RC - 2

1. Vrai.
2. Faux.

On aurait dû souligner la portion de phrase parce que (seul objet de la question VRAI-FAUX).

6 c

En 1968, la Chine et les U.S.A. ont renoué leurs relations diplomatiques, à la suite d'un match de ping-pong.

RC - 2

1. VRAI.
2. FAUX.

On aurait dû souligner en 1968 ou, mieux, commencer la phrase par « C'est en 1968 que... ».

6 d

Le phénomène électrique est associé à une libération d'électrons.

RC - 2

1. VRAI.
2. FAUX.

On aurait dû souligner libération.

D. Six règles de rédaction sur la forme

REGLE 7

La question doit être compatible avec la consigne

Ainsi, la question suivante annonce qu'une seule solution est correcte, alors que plusieurs des aliments cités contiennent des lipides (graisses cachées dans la viande, par exemple).

Question 7 a

Lequel des aliments ci-dessous apporte des lipides ?

1. œuf
2. Viande
3. huile

Il s'agit plutôt d'un continuum où l'une des solutions proposées (la solution 3) apporte *plus* de lipides que les autres.

REGLE 8

Proposer des phrases syntaxiquement correctes

Dans la question 8a, les solutions 1 et 3 ne sont pas en accord grammatical avec le tronc (cet exemple est caricatural).

8 a

On appelle translation le mouvement que la terre effectue

1. d'une durée d'un jour (soit 24 heures),
2. en 365 jours (soit un an),
3. dont la durée est de 1 440 minutes,
4. sur elle-même autour de l'axe N-S.

RC = 2

La règle 8 est souvent violée à propos du genre et du nombre des mots, ainsi qu'à propos des accords (question 8b).

8 b

L'organe qui sécrète l'insuline est le

1. pancréas.
2. rate.
3. thyroïde.
4. hypophyse.
5. foie.

RC = 1

La présence de l'article « le » dans l'amorce élimine d'emblée les solutions 2, 3 et 4.

REGLE 9

Pas de terme vague dans l'énoncé

9 a

Quel est l'élément du sang qui transporte l'oxygène ?

1. Le fer.
2. L'hémoglobine.
3. Les globules rouges.

RC = impossible à définir.

Dans la question 9a, le terme « élément » est vague : il est pris dans le sens de « composant ». Des chimistes auraient cependant tendance à prendre *élément* dans un autre sens, celui du tableau périodique. Une telle interprétation est d'autant plus probable que l'oxygène est l'un des éléments de ce tableau. Pour lever l'ambiguïté, il faut préciser si l'on se place au niveau cellulaire, moléculaire, atomique,...

REGLE 10

Eviter les formes négatives (syntaxiques et sémantiques), a fortiori proscrire leur accumulation

La présence de négations complique le traitement que l'étudiant doit opérer sur le texte. Rappelons que deux négations s'annulent et que leur accumulation transforme la question en un problème de logique. Dans la question 10a, l'accumulation des négations est outrancière ; rappelons que la négation n'est pas toujours exprimée par *ne pas* ou *non* ; elle est sémantiquement présente dans des mots tels que : *retrait, absence, interdiction, cessation, manque, nier, supprimer, empêcher, éviter, impossible*, etc.

10 a

N'est-il pas faux de nier que l'absence d'arrêt à un signal « STOP » n'entraîne pas automatiquement le retrait du permis de conduire ?

1. Oui.
2. Non.

La formulation affirmative est préférable. Si l'on ne peut éviter la négation, il importe de bien la mettre en évidence (on souligne, on écrit en gras, etc.).

Dans la question 10b, on attire l'attention du lecteur sur une difficulté présente dans certaines langues et non dans d'autres (par exemple en anglais, la négation est renforcée par le verbe utilisé négativement : « No, I do not »).

10 b

Sur les autoroutes d'Allemagne de l'Ouest, les automobilistes ne peuvent pas dépasser 130 km/heure.

1. Oui.
2. Non.

Si l'étudiant pense que l'affirmation est correcte, il aura tendance à répondre « Non, ils ne peuvent pas » (à cause de *ne pas*). Il choisira alors 2 (non) en pensant « Oui, cette affirmation est correcte ». Il suffit de remplacer *oui/non* par *vrai/faux* pour supprimer l'ambiguïté.

La question 10c met l'accent sur une difficulté particulière : l'interprétation de la structure « n'est pas... parce que... ».

10 c

Le cheval n'est pas un plantigrade parce qu'il marche sur un doigt.

1. Vrai.
2. Faux.

On peut interpréter la phrase de la question 10c de deux manières différentes :

« Le cheval est un plantigrade, *mais ce n'est pas parce qu'il marche sur un doigt* ».

« Le cheval n'est pas un plantigrade, pour la *bonne raison* qu'il marche sur un doigt ».

Cette ambiguïté doit évidemment être évitée.

REGLE 11

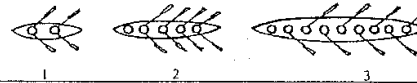
Séparer informations et questions

Dans la question 11a, la question et les informations sont trop entremêlées.

NON PAS

11 a

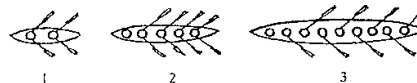
Lequel des trois dessins ci-dessous représente une embarcation armée en pointe (le rameur tire l'aviron des deux mains) avec barreur (l'équipier qui tient la barre du gouvernail) ? Dans une embarcation armée en couple, il y a deux avirons par rameur. La définition d'aviron est « rame élargie à un bout, en forme de pelle ». Le mot « aviron » vient de l'ancien français « viron », qui signifiait « tour ».



MAIS

11 b

Dans une embarcation armée
 . en *pointe*, le rameur tire un aviron des deux mains.
 . en *couple*, chaque rameur a deux avirons.
 La définition d'aviron est « rame élargie à un bout, en forme de pelle ». Le barreur est l'équipier qui tient la barre du gouvernail.
 Lequel des trois dessins ci-dessous représente une embarcation armée en pointe avec barreur ?



REGLE 12

Regrouper dans l'amorce les éléments communs aux solutions proposées

Cette règle, pourtant banale et portant sur un aspect superficiel, est souvent transgressée.

NON PAS

12 a

La température interne du corps humain :

1. ne peut jamais dépasser 37° centigrades sous peine de mort.
2. ne peut jamais dépasser 40° centigrades sous peine de mort.
3. ne peut jamais dépasser 42° centigrades sous peine de mort.

MAIS

12 b

L'homme meurt si sa température interne dépasse :

1. 37°
2. 40°
3. 42°

RC = 3

NON PAS

12 c

La température de solidification d'une matière est :

1. inférieure à la température de fusion de cette matière.
2. égale à la température de fusion de cette matière.
3. supérieure à la température de fusion de cette matière.

MAIS

12 d

La température de solidification d'une matière est :

1. inférieure
2. égale
3. supérieure

à sa température de fusion.

RC = 2

D) Huit règles de rédaction des solutions proposées

Les huit règles qui suivent portent sur quatre points envisagés d'une part dans leur aspect formel et d'autre part dans leurs liens avec le contenu. On a ainsi quatre paires de règles :

ASPECT FORMEL

Règle 13 : indépendance syntaxique des solutions.

Règle 15 : égalité de mots communs à la solution et à l'amorce.

Règle 17 : même longueur pour toutes les solutions.

Règle 19 : même degré de généralité des mots déterminants (quantificateurs, adverbes de temps, etc.).

ASPECT DU CONTENU

Règle 14 : indépendance sémantique des solutions.

Règle 16 : égalité de vraisemblance entre solutions.

Règle 18 : même complexité de toutes les solutions.

Règle 20 : même degré de technicité du vocabulaire dans toutes les solutions.

REGLE 13

Indépendance syntaxique

13

- Le muscle du bras dont la contraction permet d'effectuer le mouvement représenté ci-contre est :
1. le deltoïde.
 2. le triceps.
 3. au contraire, le biceps.
 4. en plus, le cubitus de la main.



Dans cet exemple caricatural, les expressions « au contraire », « en plus », lient ces solutions à d'autres : c'est un défaut grave.

REGLE 14

Les solutions proposées doivent être sémantiquement indépendantes les unes des autres

Dans la question 14, les solutions proposées sont *emboîtées* les unes dans les autres.

- 14
- Une solution d'eau saturée en sel de cuisine peut descendre *sans geler* jusqu'à une température de
1. 0° centigrade.
 2. -2° centigrade.
 3. -4° centigrade.
 4. -6° centigrade.

Choisir la solution 4 équivaut à accepter toutes les autres (si l'eau peut descendre jusqu'à -6°, elle peut aussi descendre à -4°, -2°, 0°).

Pour lever l'ambiguïté, il suffit de préciser «... peut descendre *au maximum jusqu'à...*».

REGLE 15

Mêmes mots communs avec l'amorce

La question 15 contrevient à cette règle.

- 15
- Quel est le nom de l'organisme de police international ?
1. C.I.A.
 2. F.B.I.
 3. Interpol.
 4. S.A.S.

Dans cette question 15, les mêmes racines (*inter* et *pol*) se retrouvent dans l'amorce *et* dans une des solutions, ce qui donne à celle-ci une bien plus grande attractivité.

REGLE 16

Même vraisemblance

- 16
- Quel groupe sanguin est « receveur universel » ?
1. A
 2. B
 3. AB
 4. O
 5. AO
 6. BO

Cet exemple est caricatural : pourquoi créer de faux groupes sanguins dans ce cas ?

REGLE 17

Même complexité.

La solution correcte ne doit pas être systématiquement plus longue que les autres

- 17 a
- Le mot « spinaker » désigne :
1. un chien écossais.
 2. une voile triangulaire de grande surface, légère et très creuse, que les yachts envoient dans la marche au vent arrière.
 3. un chant tyrolien.
 4. un instrument de musique.

RC = 2

Il est difficile pour le rédacteur d'éviter cette asymétrie. Trouver des solutions *indiscutablement* fausses est aisé, il est souvent beaucoup plus difficile d'en fournir d'*irréprochablement* correctes. Alors, on précise, on nuance. Mais l'élève n'est pas dupe ! Si on ne sait pas rendre concise la solution correcte, il s'indique d'allonger les solutions incorrectes.

17 b

En statistique, la formule de l'écart type (σ), calculé à partir de notes X de moyenne X est :

$$1. \sigma = \sqrt{\frac{\sum X^2}{N}}$$

$$2. \sigma = \sqrt{\frac{\sum (X-\bar{X})^2}{N}} \text{ pour une population et } S = \sqrt{\frac{\sum (X-\bar{X})^2}{N-1}} \text{ pour un échantillon}$$

$$3. \sigma = \sqrt{\sum (X-\bar{X})^2}$$

RC = 2

Dans la question 17b, il suffit d'introduire la contrainte « population » dans l'amorce pour ramener la solution 2 à une formule unique (la première) : « En statistique, la formule de l'écart-type (σ) d'une distribution de notes X (de moyenne \bar{X}) constituant la population envisagée est : ».

REGLE 18

Même complexité

Dans la question 18, la solution correcte ne doit pas apparaître comme plus complète que les autres.

18

Un incendie se déclare dans une cuisine. L'occupant n'arrive pas à le maîtriser. Que doit-il faire ?

1. Couper l'arrivée du gaz au compteur.
2. Téléphoner aux pompiers.
3. Evacuer les enfants.
4. 1 et 2.
5. 2 et 3.
6. 1, 2 et 3.

RC = 6

La question 18 comprend « des précautions à prendre » et « des opérations techniques à effectuer ». Dans ce genre de contenu, le constructeur des questions est tenté de construire la solution correcte en accumulant simplement l'ensemble des précautions ou des opérations. Si cette tendance est systématique chez le constructeur, alors il suffit à l'étudiant de rechercher la solution la plus complète.

REGLE 19

Même niveau de généralité dans les indicateurs (temps, modificateurs ...) !

Les termes *tous*, *toujours*, *jamais*, *aucun*, etc. sont absolus et catégoriques. Les étudiants s'en méfient et évitent de choisir les solutions où ces termes figurent. Par contre, ils préfèrent les solutions contenant des termes tels que *certains*, *parfois*, *il peut arriver que*, etc.

19 a

Dans une distribution normale (courbe de Gauss), la moyenne

1. est toujours supérieure à l'écart type.
2. est toujours inférieure à l'écart type.
3. est parfois inférieure à l'écart type.

RC = 3

19 b

Le terme *épicea* désigne :

1. tous les arbres résineux.
2. tous les pins.
3. certains types de pins.

RC = 3

Dans les questions 19a et 19b, la solution 3 est rendue plus attractive que les deux autres.

REGLE 20

Même degré de technicité du vocabulaire dans toutes les solutions proposées

Les étudiants peu compétents ont tendance à éviter les solutions comportant des termes techniques. Dans la question 20, les solutions 1, 2 et 4 contiennent, pour l'étudiant ignorant, des mots plus difficiles que ceux de l'énoncé. La solution 3 est donc avantagée.

20

Dans l'Islam chiite, un ayatollah est

1. un imam.
2. un uléma.
3. une autorité religieuse.
4. un derviche.

RC = 3

F. L'habileté à répondre à des QCM

Certains étudiants même inexpérimentés sur le fond, acquièrent une grande habileté à détecter la réponse correcte d'une QCM, et notamment à partir des erreurs de rédaction commises par l'auteur de la question.

1. Définition

Cette habileté est appelée, en anglais, *test wiseness*, que De Landsheere (1979) définit comme la *qualité de celui qui est aguerri aux tests*.

L'antonyme de *test wiseness* pose moins de problèmes de traduction : Thomdike (1971, p. 364) parle de *test naiveté* (en anglais).

Gibb (1964) définit la *test wiseness* comme « la capacité d'un étudiant à répondre correctement à un test fait de QCM qui contiennent des repères formels et, par conséquent, la capacité à obtenir des points sans connaître la matière testée ».

Nous verrons que l'étude expérimentale de cette habileté des étudiants est un des plus sûrs moyens de découvrir ou de valider des règles de rédaction des QCM. Il suffit de *transgresser* l'une de ces règles, en créant des questions volontairement biaisées, puis d'observer le comportement des étudiants.

2. Les méthodes d'étude de cette habileté

Deux techniques expérimentales ont été utilisées.

- a) Certains auteurs ont *introduit des erreurs de rédaction* dans des QCM bien rédigées (et qui avaient été éprouvées expérimentalement). Les résultats obtenus après l'introduction de l'erreur sont comparés aux résultats obtenus avec la forme initiale. Board et Witney (1972) ont procédé de la sorte sur un test destiné à des étudiants universitaires ; Mc Morris, et al. (1972) ont fait de même sur un test destiné à des étudiants du secondaire supérieur.
- b) Le plus souvent, on propose à l'étudiant des *questions dont le contenu est fictif* (inventé de toutes pièces par l'auteur). Cette technique présente deux avantages majeurs :
 - 1° Il ne peut y avoir d'interférence entre les résultats et le degré de maîtrise de la matière chez les divers étudiants (puisque cette matière n'existe pas).

2° Le taux de choix de chaque solution devrait correspondre au hasard : par exemple, 25 % pour quatre solutions proposées. Toute *transgression* d'une règle de rédaction qui « *avantage* » une solution devrait se refléter directement dans le taux élevé de choix de celle-ci (plus de 25 % des fois). Dans les exemples qui suivent, nous marquerons d'un *astérisque (*)* la *solution avantagee*.

On trouvera dans la section G les principaux résultats obtenus au cours de diverses recherches utilisant des questions à contenu fictif. Les étudiants devaient soit choisir une solution, soit juger chaque solution proposée. Plusieurs de nos exemples seront empruntés aux trois études brièvement décrites ci-dessous.

Slakter, Kochler et Hampton (1970) ont transgressé quatre règles, chacune dans quatre QCM fictives. Les seize questions ont été mêlées à vingt-huit QCM normales et présentées à des étudiants de dix-sept ans. De plus, ces auteurs ont mesuré la *test wiseness avant et après un apprentissage spécifique* par enseignement programmé sur la façon de tirer parti de ces quatre erreurs.

Diamond et Evans (1972) ont transgressé cinq règles, chacune dans six QCM fictives. Le test de trente questions ainsi obtenu a été présenté à des élèves de douze ans.

Strang (1977) a introduit quatre QCM *fictives* parmi huit QCM normales, au niveau universitaire. Les étudiants étaient avertis que chaque solution devait être considérée en soi, car plusieurs pouvaient être correctes. Les étudiants pouvaient donc choisir plusieurs solutions par question.

Il est impossible de relater ici toutes les expériences faites dans ce domaine. Nous nous bornerons aux plus révélatrices d'entre elles à la fois par les résultats obtenus et par les moyens (souvent ingénieux) mis en œuvre pour les recueillir.

G. Résultats d'études expérimentales de l'habileté à répondre à des QCM

1. Transgression de la règle 8

Dans ces QCM l'article *an* (en anglais) et le pronom *il* (dans notre traduction française) fournit un indice facilitateur.

Exemple de question à contenu fictif (adapté de Diamond et Evans, 1972).

La dernière amitié de Harlan Stone est un

1. satire politique ?
- *2. récit autobiographique ?
3. histoire de science-fiction ?
4. courte biographie ?

Résultats

La solution avantagee a recueilli 35 % des choix (Diamond et Evans), et non les 25 % attendus. L'effet de l'article (*a* ou *an*) a aussi été mis en évidence par les études de Dunn et Goldstein (1959), et Mc Morris et al. (1972).

2. Transgression de la règle 14

Les solutions ne sont pas indépendantes.

a) Exemple de question fictive

La question ci-dessous est due à Diamond et Evans (1972). Si la solution 1 est correcte, les autres ne le sont pas.

La population de Frankton est inférieure à

- *1. 50 mille habitants
2. 60 mille habitants
3. 70 mille habitants
4. 80 mille habitants

b) Autre exemple de question fictive

La question ci-dessous est due à Slakter et al. (1972). Si la solution 2 est correcte, les solutions 1 et 4 le sont aussi. Seule la solution 3 est indépendante des autres : elle est donc avantagee.

Quand on ajoute des cristaux de Bestor à l'eau,

1. De la chaleur se dégage.
2. La température de la solution s'élève.
- *3. La solution devient bleue.
4. Le récipient devient plus chaud.

Résultats.

Au lieu des 25 % théoriquement attendus, la solution avantagee a reçu des pourcentages de choix égaux à

- 45 % dans l'expérience de Diamond et Evans (1972) ;
- 52 % avant l'apprentissage et 79 % après apprentissage dans l'expérience de Slakter et al. (1972).

3. Transgression de la règle 15

Un mot (ou une phrase) est répétée dans l'amorce et dans les solutions proposées.

Exemple de question fictive (due à Marshall et Hales, 1971)

Le Parti national augustin a son quartier général à

1. Camden, dans le New Jersey.
- *2. St Augustin, en Floride.
3. Palo Alto, en Californie.
4. Dallas, au Texas.

Résultats

- La solution avantagee a reçu
- 76 % des choix (Diamond et Evans, 1972)
- 66 % des choix avant et 82 % après apprentissage (Slakter et al. 1972).

Des résultats semblables ont été obtenus par les études de Dunn et Goldstein (1959) et Mc Morris et al. (1972).

4. Transgression de la règle 16

Les solutions sont toutes absurdes, sauf une (qui, par le fait même est avantagée).

Exemple de question fictive (construite par nous afin d'illustrer le procédé)

Le vent du désert de Zubi est fait

1. De petits insectes.
2. De cristaux de cuivre.
- *3. De cristaux de glace.
4. D'aiguilles de pins.

Résultats

Dans une expérience de Slakter et al. utilisant ce genre de question, la solution avantagée a recueilli 84 % de choix *avant* et *après* apprentissage.

Le taux élevé de réponses correctes avant l'apprentissage et le peu d'effets de ce dernier laissent penser que la présence de distracteurs absurdes a rendu « évidentes » les questions fictives. Celles-ci ne sont donc peut être pas une bonne méthode pour étudier ce défaut de rédaction particulier.

5. Transgression de la règle 17

La solution correcte est plus longue que les distracteurs.

Exemple de question fictive (due à Marshall et Hales, 1971)

La loi Davis (XIXe siècle)

1. Fournit de l'argent aux écoles
2. Abolît une loi précédente
- *3. Interdit la fabrication, la vente, le transport ou l'utilisation de divers médicaments qui étaient alors utilisés dans des intentions illégales.
4. Augmenta le salaire des agents de l'Etat.

Résultat :

La solution avantagée a reçu 53 % des choix (Diamond et Evans, 1972). Des résultats semblables ont été obtenus par Chase (1964), Dunn et Goldstein (1959), Mc Morris et al. (1972).

L'expérience de Strang (1977) permettait d'étudier aussi l'effet de la longueur des solutions. Les solutions longues reçurent 57 % de réponses VRAI et les courtes 38 % de réponses VRAI, au lieu de 50 % théoriques.

6. Transgression de la règle 19

La présence de mots tels que *tous, toujours, jamais*, etc. dans certains distracteurs les rend peu attrayants.

Exemple de question à contenu fictif (due à Marshall et Hales, 1972)

Herman Klavemann est célèbre parce que

1. Il a développé toutes les échelles musicales utilisées en Occident.
2. Il a composé toutes les sonates pendant la période romantique.
3. Il a traduit tous les classiques russes en anglais.
- *4. Il a inventé l'épingle de sûreté.

Résultats

Dans l'expérience (due à Diamond et al.), la solution avantagée a recueilli 50 % des choix, au lieu des 25 % théoriquement attendus.

Dans l'expérience de Slakter et al., la solution avantagée a reçu 41 % des choix avant apprentissage (de « test wiseness ») et 75 % des choix après cet apprentissage.

7. Transgression de la règle 20

Le vocabulaire n'a pas le même degré de technicité dans toutes les solutions.

Exemple de question fictive (due à Strang 1977).

Dans la théorie de J. Havighurst

1. Les qualités *isofitales* sont importantes.
2. Le jeu *acropastique* joue un grand rôle dans le développement de la personnalité morale de l'enfant.
- *3. Les peurs des enfants sont spécialement étudiées.
- *4. On considère la personnalité comme la dimension la moins bien comprise de l'ensemble de l'être humain.

Cette question permet d'étudier à la fois les effets de la règle 20 et ceux de la règle 17 (les solutions ne comportant pas de termes techniques sont « avantagées » ; les solutions 1 et 3 sont courtes, les solutions 2 et 4 sont longues).

On remarquera que même les termes techniques sont fictifs.

Dans l'expérience de Strang, l'étudiant doit juger chaque solution proposée.

Les solutions avantageées sont jugées vraies dans 60 % des cas (contre 50 % en théorie) alors que les « désavantageées » sont jugées vraies dans 35 % des cas (contre 50 % en théorie).

Dans la question ci-dessous, Strang a construit ces solutions pour présenter :

(en 1), aucun terme technique (donc tous termes courants).

(en 2), un terme technique courant (*).

(en 3), un terme technique relativement courant (*).

(en 4), un terme technique peu courant (*).

(*) que nous avons écrit en italique

Heinz Werner

1. a dressé le graphique du développement physique des enfants.
2. a mis au point une technique d'analyse des chromosomes.
3. a insisté sur l'importance de l'hétéronomie.
4. croyait que l'aloplasie était héréditaire.

Avant de construire les questions, l'auteur a établi expérimentalement à quelle catégorie (familier, vaguement familier, non familier) appartenait chaque terme technique pour les élèves considérés.

Des « juges » (étudiants du même âge) ont été invités à classer des termes techniques dans les trois catégories, à partir des définitions suivantes :

familier = je peux le définir grossièrement.

vaguement familier = je ne peux le définir, mais je l'ai déjà rencontré auparavant.

non familier = je ne me rappelle pas l'avoir rencontré.

Pour la question ci-dessus, les étudiants devaient de nouveau juger chaque solution proposée. Pour ces quatre types de solutions, les taux de jugement « VRAI » ont été, respectivement de 55 %, 41 %, 38 % et de 25 % (au lieu des 50 % théoriques).

8. L'ordre des solutions correctes dans un test

Les solutions correctes des questions successives forment une séquence systématique facilement repérable.

Rabinowitz (1970) a observé que les réponses successives aux diverses questions d'une épreuve ne sont pas entièrement aléatoires, même quand les étudiants sont entièrement ignorants. Par exemple, ils ont tendance à épuiser l'ensemble des possibilités avant de recommencer un « cycle ».

Ces observations sont confirmées par Wood (1977, p. 232) : il est très rare de rencontrer des sujets notant la même solution pour toutes les questions.

Wilbur (1970) a systématiquement fait varier la position de la solution correcte de 20 questions sur les 80 qu'en comptait un test. Dans la forme 1 du test, les 20 QCM avaient 1 comme solution correcte ; dans la forme 2, leur solution correcte était en deuxième position, et ainsi de suite. Les résultats ne montrent aucune influence de l'ordre, mais ils paraissent peu déterminants, car cet ordre systématique ne concernait qu'une question sur quatre, et par conséquent, n'a sans doute pas été perçu. Il est regrettable que l'auteur n'ait pas interrogé les étudiants à ce sujet.

9. Observations générales

D'autres observations, plus générales, ont été faites à l'occasion de ce type d'études.

1. Les infractions aux règles 12, 17 et 14 ont essentiellement pour résultat une augmentation de la facilité de la question. Leur influence sur la validité et la fidélité s'avère quasi nulle (Mc Morris, Brown et al., 1972), ce qui est surprenant. Board et Whitney observent, eux, une nette diminution de la validité et de la fidélité en cas d'infraction systématique.
2. La relation entre l'intelligence générale et l'habileté à tirer parti des défauts dans le respect des règles 12, 17, 14 semble faible (Diamond et Evans, 1972). On peut penser que même les moins intelligents sont capables de déceler des défauts aussi grossiers ! Certains étudiants sont plus sensibles à certains défauts de construction qu'à d'autres.
3. Les étudiants sont conscients de ces défauts (aux règles 12, 17, 14, 15 et 19) comme en témoignent leurs justifications a posteriori (Diamond et Evans, 1972, qui s'adressent à des étudiants de 12 ans ; Millman et al., 1965, qui s'adressent à des étudiants du secondaire).
4. Il est possible de « progresser » dans l'habileté à tirer parti de certains défauts, en s'entraînant au moyen de cours programmés appropriés. Néanmoins, ces progrès ne semblent pas se maintenir pour tous les types de défauts de rédaction (Slakter et al., 1970).
5. Selon Mc Morris et al. (1972), la méthode qui consiste à modifier des épreuves authentiques permettrait de mesurer l'habileté (test wiseness) chez des groupes d'étudiants, mais non chez des individus particuliers. D'autres recherches sur le même sujet ont été menées par Crehan, Koelher et Slakter (1974, Rowley (1974), Nilsson et Wedman (1978), ainsi que Jangma et Pound (1977).

CHAPITRE 4
**LES CONSIGNES
NON CLASSIQUES**

Introduction

- A. Les variations des modes de réponses.**
1. La réponse forcée.
 2. La consigne « n réponses ».
 3. La consigne « autant de réponses que vous voulez ».
 4. La consigne « une réponse par solution ».
- B. Discussion sur les consignes à réponses multiples.**
- C. La consigne « solutions générales » (Codes G).**
1. « Aucune » (k+1)
 2. « Toutes » (k+2)
 3. « Données insuffisantes » (k+3)
 4. « Absurdités dans l'énoncé » (k+4).
- D. Discussion sur le principe des solutions générales.**
1. Une règle fondamentale.
 2. Un exemple d'utilisation systématique.
 3. Critique des solutions générales.
 4. Solutions générales et questions VRAI-FAUX.
 5. D'autres solutions générales.
- E. Exemple d'épreuves recourant aux solutions générales.**
1. Géométrie.
 2. Géographie.
 3. Alphabet.
- F. Les codes de chaque QCM**

Introduction

Dans un premier temps, nous repartirons de la taxonomie des consignes Q et R, pour explorer les variations des modes de réponses (R).

Dans un second temps, nous examinerons une autre composante possible de la consigne : les codes G précisant l'existence (ou non) de solutions générales dans les questions.

Cette consigne, assez nouvelle, sera illustrée par un test entier qui y recourt.

A. Les variations des modes de réponses

1. La réponse forcée ou l'abstention interdite (le code R = 0).

Ce type de question se présente dans les situations de détection : si le signal est passé et si l'étudiant n'a pas réagi, le signal est considéré comme « non repéré ». Voici un exemple de cette situation dans le domaine de l'orthographe.

Consigne QR = 00 Q0 = Une seule (ou aucune) des solutions proposées (n°) est correcte.

R0 = Fournissez une réponse pour chaque solution proposée, l'abstention constituant une réponse.

Dans le texte ci-dessous, certains mots sont numérotés et soulignés. Barrez ceux qui comportent une faute d'orthographe.

Au moindre bruit suspet¹, le lièvre poltron²
détalle³ à toute vitesse à travers chant⁴
⁵ ⁶ ⁷ ⁸

Cette modalité de réponse (Code R = 0) est impossible avec des consignes où le code Q est différent de 0. Donc les consignes où les codes QR seraient 10 ou 20 ou 30 ou 40 ou 50 n'existent pas.

Nous avons vu (chapitre 2) les très nombreuses consignes où il fallait donner une seule réponse (codes QR = 01, 11, 21, 31, 41 et 51).

2. La consigne « n réponses » ou le code R = 2 : Fournissez × réponses.

Consigne QR = 22 Q2 = 3 solutions proposées sont correctes.
R2 = Fournissez 3 réponses.

Quels sont les trois pays Africains qui, au début des années 60, ont constitué une entente cordiale, sous l'égide de leurs chefs d'Etats respectifs : Sékou Touré, Kwamé Nkruma, Modibo Keita ?

- | | |
|------------------|----------------|
| 1. Côte-d'Ivoire | 5. Mali |
| 2. Guinée | 6. Sénégal |
| 3. Togo | 7. Haute Volta |
| 4. Ghana | 8. Niger |

RC = 2, 4, 5.

L'inconvénient de ce mode de réponse est que l'étudiant peut être très sûr de l'une de ses trois réponses, moins sûr pour une deuxième et pas du tout pour la troisième (qu'il a choisie au hasard). Le correcteur ignorera ces différences entre réponses, notamment parce que le nombre a été imposé (sans quoi notre étudiant n'en aurait sans doute fourni que deux). De plus, on devrait préciser quelle décision prendra le correcteur si l'étudiant fournit moins ou plus de réponses que le nombre exigé.

Les consignes QR = 02 et QR = 12 n'existent pas : on ne peut demander de choisir plus de solutions correctes qu'on en annonce.

3. Le code R = 3 : Fournissez autant de réponses que vous voulez.

Ici, le nombre de réponses à fournir n'est pas imposé. Il peut varier entre 1 et k (le nombre de solutions proposées). Cette modalité de la consigne pose des problèmes pratiques. Il est en effet impossible de concevoir une grille de réponses, puisque l'on ignore le nombre de réponses que donnera l'étudiant. Des complications surgissent aussi quand on veut encoder les réponses pour les faire corriger par un programme d'ordinateur.

a) Le nombre de solutions correctes étant annoncé (QR = 23)

Une variante intéressante de cette consigne est inspirée par le modèle 2 de l'activité mentale. Rappelons que pour ce modèle, l'étudiant choisit la réponse après avoir éliminé les solutions qu'il sait incorrectes.

Afin de mesurer cette « connaissance partielle », Coombs, Milholland et Woormer (1956) ont proposé pour les QCM les plus classiques (une seule solution correcte), la consigne « *Éliminez les solutions incorrectes* » (rejet).

Consigne : QR = 23 Q2 = 3 solutions proposées sont incorrectes.
R3 = Éliminez-les. Fournissez autant de réponses que vous voulez.

Ex :

Les insectes portent les pattes sur :

1. La tête.
2. Le thorax.
3. L'abdomen.
4. Les antennes.

Éliminez les solutions incorrectes !

RC = 1, 3 et 4.

A raison d'un point par rejet pertinent, la *réponse parfaite* reçoit ici 3 points, soit $k-1$. Pour une question, la gamme des scores positifs possibles va donc de 1 à 3, ce qui est plus nuancé que l'alternative habituelle 0 ou 1. Cette procédure a d'ailleurs été développée pour mieux évaluer l'*information partielle* (voir aussi Collet, 1971).

Si l'étudiant élimine la solution correcte (rejet non pertinent), on lui retirera ici 3 points, soit une *pénalisation de $k-1$* . Par conséquent, un étudiant qui s'abstient de répondre obtient 0, de même qu'un étudiant qui élimine toutes les solutions : ce dernier gagne $(k-1)$ fois 1 point, mais perd une fois $(k-1)$ points.

Voici les croix faites par huit étudiants et les huit scores obtenus.

	Etudiants							
	A	B	C	D	E	F	G	H
1. La tête	1	1	X	1	X	X	X	1
2. Le thorax	2	2	2	2	X	X	X	X
3. L'abdomen	X	X	X	3	X	3	3	3
4. Les antennes	4	X	X	3	X	X	4	4

Score à la question	1	2	3	0	0	-1	-2	-3
---------------------	---	---	---	---	---	----	----	----

b) Le nombre de solutions correctes n'est pas annoncé (QR = 33)

Quand le nombre de solutions correctes (NSC) est supérieur à 1, le score maximum vaut alors $k - \text{NSC}$. En cas d'élimination (non pertinente) d'une des NSC solutions correctes, la pénalisation vaudra $\frac{k - \text{NSC}}{\text{NSC}}$.

Un étudiant qui élimine toutes les solutions obtiendra un score égal à 0. Appliquons ces formules à la question suivante où le nombre de solutions à rejeter n'est pas annoncé à l'étudiant.

Consigne QR = 33 Q3 : Une ou plusieurs solutions sont correctes.
R3 = Fournissez autant de réponses que vous voulez.

Biffer le (ou les) Etats qui ne sont pas traversés par un fleuve nommé Colorado.

1. Massachusetts.
2. Arizona.
3. Californie.
4. Texas.
5. Argentine.

Le patron de réponses $\cancel{1} \ 2 \ 3 \ 4 \ 5$ est noté + 2 points (maximum)

Le patron de réponses $1 \ 2 \ \cancel{3} \ 4 \ 5$ est noté $-\frac{5-3}{3} = -\frac{2}{3}$

Le patron de réponses $1 \ \cancel{2} \ \cancel{3} \ 4 \ 5$ est noté - 2 points (minimum).

c) Il pourrait n'y avoir aucune solution correcte (QR = 43)

Consigne QR = 43 Q4 = Une, plusieurs ou aucune des solutions proposées (n') est (son) correcte(s).
R3 = Fournissez autant de réponses que vous voulez.

Parmi les personnes suivantes qui ont reçu un Prix Nobel, lesquelles ont obtenu le Prix Nobel de la Paix ?

- | | |
|---------------------------|-----------------------|
| 1. Henri Dunant | 5. Martin Luther King |
| 2. Albert Schweitzer | 6. Jacques Monod |
| 3. Jean-Paul Sartre | 7. Marie Curie |
| 4. Alexandre Soljenitsyne | 8. Winston Churchill |

RC = 1, 2, 5

Cette QCM peut être utilisée avec la consigne QR = 33. L'utiliser avec la consigne QR = 43 la rend cependant plus difficile.

d) Solutions pondérées sur un continuum de qualité (QR = 53)

Dans le cas où les solutions se situent sur un continuum de qualité, l'étudiant recevra un score égal à la somme des valeurs attachés à chaque solution choisie.

Consigne QR = 53 Q5 = La qualité des solutions proposées varie sur un continuum de -4 à +4.

R3 = Fournissez une ou plusieurs réponses.

Exemple (dû à Perée)

La fonction $y = \sqrt{\frac{x^3}{x-1}}$	Conséquences
1. est définie sur $] -\infty, 1[\cup] 1, +\infty [$	-3
2. est égale à 2 pour $X = -1$	-3
3. a pour dérivées $y' = \frac{1}{2} \sqrt{\frac{x^3}{x-1}} - 3(2x-3)$	+4
4. a pour primitive $y = \frac{2}{9} (1+x^3)^3 + C$	-4
5. admet la droite $y = x$ comme asymptote	+4
6. est continue en $x = 0$	+2
7. est impaire	-3
8. admet un minimum en $x = \frac{3}{2}$	+3

Ce type de questions élargit la gamme des notations possibles et permet de discriminer les étudiants. Dans l'exemple ci-dessus, les scores peuvent aller de +13 points à -13 points. Ainsi l'élève qui aurait choisi les solutions 1, 3 et 6 obtiendrait : $-3+4+2 = 3$ points, sur un maximum possible de 13 points.

Ce type de questions demande à l'étudiant un travail aussi considérable qu'en cas de questionnaire « Vrai-Faux généralisé ». L'exploitation statistique qu'on peut en faire est cependant plus fruste, car on ne distingue pas ici, entre le rejet (FAUX) et l'omission, ce qui est le cas en Vrai-Faux généralisé (voir ci-après).

4. Le code R = 4 : Fournir une réponse pour chaque solution proposée

Cette modalité consiste à demander de :

- choisir les solutions correctes (en les entourant, ou en les annotant du signe +, ou du signe V).
- éliminer les solutions incorrectes (en les biffant, ou en les annotant du signe -, ou du signe F).

C'est le *jugement multiple* (Noizet et Fabre, 1975).

Cette consigne considère chaque solution comme une question en soi, à laquelle on doit répondre par VRAI ou FAUX.

Donc, chaque sous-question doit porter un numéro, alors que l'amorce n'en a pas. Une question dont la consigne est QR = 44 et qui a k solutions se transforme (pour le codage et la correction) en k questions dont le code de consigne est QR = 11.

Consigne QR = 44

Pour chacune des solutions suivantes,

Q1 = une seule des (deux) solutions proposées (VRAI ou FAUX) est correcte.

R1 = Fournissez une réponse (à chaque solution).

Des substances minérales peuvent être absorbées à partir du milieu extérieur par :

1. les prokaryotes,
2. les plantes
3. les animaux,
4. les virus.

Lorsque toutes les QCM d'un test répondent à cette même consigne, on peut parler de VRAI-FAUX généralisé ou VFG (en anglais multiple true false).

La question ci-dessus pourrait aussi bien être utilisée avec les consignes

QR = 04 (une seule ou aucune solution correcte)

14 (une seule solution correcte)

24 (x solutions correctes)

34 (une ou plusieurs solutions correctes)

et 54 (la qualité des solutions varie sur un continuum)

B. Discussion sur les consignes à réponses multiples.

La consigne « une, aucune, plusieurs ou toutes les solutions sont correctes » a été remarquablement défendue par J. Halleux-Hendrick.

Ce type de consigne convient bien aux questions portant sur des raisonnements, afin de ne pas « fausser l'optique de l'étudiant en lui laissant croire qu'à chaque problème, qu'il soit posé par un professeur ou par la vie, il n'existe qu'une seule et unique solution valable » (Halleux, 1969).

Voici une des questions (consigne QR = 33) qu'elle posait :

Question : Effectuer $(x + 5)^2 - (x - 5)^2$
Quelle(s) solution(s) présente(nt) des raisonnements corrects ?

Solutions proposées :

1. $(x+5)^2 - (x-5)^2$ = $x^2 + 10x + 25 - x^2 + 10x - 25$ = $20x$	4. $(x+5)^2 - (x-5)^2$ = $x^2 + 5^2 - x^2 + 10x - 5^2$ = $20x$
2. $(x+5)^2 - (x-5)^2$ = $x^2 + 25 - x^2 + 25$ = 50	5. $(x+5)^2 - (x-5)^2$ = $(x+5-x+5)(x+5+x-5)$ = $10 \cdot 2x$ = $20x$
3. $(x+5)^2 - (x-5)^2$ = $x^2 + 10x + 25 - x^2 + 10x - 25$ = $20x$	6. aucune des 5 solutions ci-dessus n'est correcte.

Les solutions 3, 4 et 5 sont correctes.

Toutes les trois aboutissent au même résultat (20x). Halleux est consciente de ce problème :

« Notre expérience a mis en évidence une certaine attitude des élèves, attitude dont il nous faut tenir compte dans la construction des questions à choix multiple. Il s'agit de l'influence des *réponses finales simples et identiques*. Nous entendons par réponse finale simple, la conclusion terminale d'un raisonnement qui peut s'écrire sous la forme d'un seul mot ou d'un nombre.

Nous signalons à titre documentaire, à ce propos, (deux) exercices... dans lesquels nous avons relevé une moyenne de 30 % de jugements semblables suscités par les mêmes réponses finales, indépendamment d'un raisonnement juste ou erroné. Ceci trahit une recherche de facilité de la part des élèves, un refus de refaire tout le raisonnement et une tendance à conclure que des réponses finales identiques doivent automatiquement entraîner la même prise de position.

Donc, lors de la construction des exercices, nous devons éviter de donner systématiquement aux solutions exactes une même réponse finale et aux solutions fausses une ou plusieurs autres réponses finales différentes de la précédente. Auquel cas, en effet, nous ne développerions pas l'esprit de discrimination des élèves.

Ce problème ne se pose pas dans tous les domaines. Par exemple, en géométrie, dans les applications se rattachant à la similitude, la thèse est chaque fois bien définie et toutes les démonstrations proposées doivent se terminer par une conclusion identique à la thèse. Les élèves ne se laissent donc pas influencer par les réponses finales qui inévitablement sont semblables. » (p. 114).

Halleux (1969a et 1969b) souligne l'intérêt de cette consigne (aucune, une, plusieurs, toutes) en partant du raisonnement suivant :

- « - On déclare fréquemment que les questions ouvertes, où l'étudiant doit construire sa réponse, sont toujours plus pertinentes pour mesurer la maîtrise d'un raisonnement » (1969a).
- Or on observe que « les étudiants fournissent le plus souvent les raisonnements et les démarches correctes telles qu'elles ont été vues en classe, reproduisant ainsi mécaniquement l'enseignement » (1969a).
- Ainsi, dans une question où il s'agit de « former un ensemble produit à partir de deux ensembles », l'auteur constate que « dans les interrogations de création, 80 % des réponses sont parfaites, mais identiques à l'exemple proposé en classe... Le professeur obtient certes de bons résultats ... mais au détriment du développement de la polyvalence et de l'adaptabilité de l'élève » (1969b, p. 124-125).
- Dans la question C ci-dessus, par exemple, la solution 5 met en œuvre un raisonnement peu courant chez les élèves :
 $a^2 - b^2 = (a + b)(a - b)$.
 Cette solution est bien plus difficile à comprendre que les deux autres solutions correctes » (1969a).

A propos d'un autre exercice, Halleux constate que seulement 18 % des élèves ont repéré toutes les solutions correctes ; 71 % des élèves n'ont admis que la solution vue en classe.

Elle conclut « ... ce sont les réponses aux QCM et non aux questions ouvertes qui nous ont permis de déceler un oubli du professeur ou une lacune dans son enseignement. Bien entendu, si la QCM n'avait comporté qu'une solution vraie, semblable à l'exemple traité en classe, elle n'aurait rien pu mettre en évidence » (p.125).

C. La consigne « solutions générales » (code G).

Par souci de clarté et d'économie, on utilise souvent la même consigne pour toutes les QCM d'une épreuve donnée. Cette consigne est alors annoncée au début de l'épreuve ; elle n'est plus rappelée par la suite.

Des indications générales peuvent parfois être aussi données pour les solutions proposées. En effet, certains types de réponses peuvent convenir à de nombreuses QCM, quel que soit leur contenu. Rappelons, par exemple, les questions du modèle « ABDO », qui comptent quatre solutions : A, B, les deux (D) et aucune des deux (0).

Ces deux dernière solutions peuvent se retrouver telles quelles dans bien des QCM différentes. On peut, dès lors, les mettre elles aussi en évidence dans la consigne. C'est le principe des solutions générales.

Elles sont valables pour toutes les questions que comporte l'épreuve.

Il faut, dans ce cas, leur attribuer un code général qui permette aux élèves de les choisir sans risque de confusion.

Par exemple, on peut convenir, avec les étudiants, que le code 9 signifiera toujours « Aucune des solutions proposées n'est correcte », le code 8 « Toutes les solutions proposées sont correctes », etc. Nous suggérons une autre façon de coder ces solutions générales. Si k représente le nombre de solutions proposées dans la QCM, on peut convenir que le nombre $k+1$ codera la solution générale « Aucune des k solutions présentées n'est correcte », le nombre $k+2$ « Toutes les k solutions présentées sont correctes, etc. Cette façon de procéder augmente la difficulté de l'examen. Elle peut devenir insécurisante. Aussi ne l'avons-nous appliquée, jusqu'à présent, que dans l'enseignement supérieur. Il reste à explorer expérimentalement dans quelle mesure l'utilisation sera possible avec des sujets plus jeunes ou doués d'aptitudes moindres.

- Ci-après, nous décrirons une consigne par trois codes :
- le code Q concernant les questions
 - le code G
 - le code R concernant les réponses.

1. La solution générale k+1 : « Rejet »

Pour de jeunes élèves, k sera remplacé dans la consigne par « nombre de solutions proposées ».

On crée ainsi (voir question b, ci-dessous) une situation de réponse différente de la consigne habituelle : « Il y a une et une seule réponse correcte » (question a ci-dessous).

Consigne donnée aux élèves au début de l'épreuve.

- a Q = 1 : Une seule solution proposée est correcte
G = 0
R = 1 : Fournissez une seule réponse (ou omettez)
- b Q = 1 : Une seule solution proposée est correcte
G = 1 : Répondez par k+1 si aucune solution proposée n'est correcte : rejet
R = 1 : Fournissez une seule réponse (ou omettez)

Question a
(consigne habituelle)

La capitale de la France est :
1. Marseille.
2. Strasbourg.
3. Lyon.
4. Lille.
5. Une autre ville.

Réponse correcte = 5

Question b
(consigne k+1)

La capitale de la Belgique est :
1. Anvers.
2. Gand.
3. Liège.
4. Mons.

Réponse correcte = 5

On recommande souvent aux étudiants, chaque fois que la question s'y prête, de répondre mentalement à la QCM avant d'examiner les solutions proposées, puis de confronter leur réponse à ces solutions. Un étudiant qui agit de la sorte et a pensé à « Paris » pour la question a verra sa réponse « confirmée » par la présence de la solution 5. Le même étudiant, répondant mentalement « Bruxelles » à la question b, ne découvrira pas une telle confirmation dans les solutions proposées.

Dans les deux cas, la réponse correcte ne figure pas parmi les solutions proposées. Mais on peut dire que dans le second cas (question b, consigne k+1), la vigilance et l'attention de l'étudiant sont davantage sollicitées ; la

performance est donc plus difficile. La QCM échappe dès lors partiellement à un reproche qu'on lui adresse fréquemment : permettre aux étudiants de fournir la réponse correcte en la « reconnaissant » parmi les solutions proposées, alors qu'ils seraient incapables de l'évoquer de mémoire.

2. La solution générale k+2 : « Toutes »

En principe, on pourrait introduire dans les consignes la solution générale « Toutes » sans proposer, en même temps, « Aucune ». Pour des raisons pratiques, nous conseillons d'utiliser les diverses solutions générales *en plus* de celles qui existent déjà, afin de garder les numéros de code k+1, k+2, k+3, etc., constants. Ainsi, quand on annonce que la consigne « k+3 » est en vigueur, cela signifie que k+2 et k+1 le sont aussi (mais pas k+4 et au-delà).

Consigne donnée aux élèves au début de l'épreuve :

- Q = 1 : Une seule solution proposée est correcte
G = 2 : Répondez par K+1 si aucune solution proposée n'est correcte
Répondez par K+2 si toutes les solutions proposées sont correctes
R = 1 : Fournissez une seule réponse (ou omettez)

Exemple

Question c

Où se trouve la ville de Namur ?
1. Au sud-est de Bruxelles.
2. Au confluent de la Sambre et de la Meuse.
3. En Wallonie.

Réponse correcte = 5 (Toutes les trois sont correctes).

On voit qu'avec le principe des solutions générales, les chances de fournir une réponse grâce au seul hasard ne sont plus de $1/k$ mais de $1/(k+g)$ où g représente le nombre de solutions générales. Dans l'exemple ci-dessus, la différence est d'importance puisque l'on passe de $1/3$ à $1/5$.

3. La solution générale k+3 : « Données insuffisantes »

Pour la plupart des QCM, on peut introduire cette solution, soit dans la question, soit sous forme de *solution générale*, avec un code (k+3, par exemple).

La consigne reprend les consignes k+1 et k+2 et ajoute :

- Q = 1 : Une seule solution proposée est correcte
 G = 3 : Répondez par K+1 si aucune solution proposée n'est correcte
 Répondez par K+2 si toutes les solutions proposées sont correctes
 Répondez par K+3 si les données de l'énoncé sont insuffisantes
 R = 1 : Fournissez une seule réponse (ou omettez)

Exemple
 Question d

Un circuit électrique a une résistance de 1 ohm. Quelle est la tension (d.d.p.) aux bornes de ce circuit ?

1. 24V
2. 110 V
3. 220 V
4. 380 V

Avec la consigne K+3, la réponse correcte est 7, car, sans connaître l'intensité (I) du courant, il est impossible de répondre à la question.

Avantages de cette consigne

Cette solution générale convient particulièrement pour tester des objectifs relevant du niveau « application » (au sens donné à ce terme dans la taxonomie de Bloom). Appliquer, c'est non seulement mettre en œuvre des lois, des principes, des formules, mais c'est aussi voir si l'on possède ou non les éléments nécessaires pour les appliquer. La solution générale k+3 permet de créer des questions à l'intersection des niveaux taxonomiques de l'application et de l'analyse.

Dangers de cette consigne

Il arrive que, pour des QCM dont la réponse correcte est k+3 (données manquantes), des élèves peu habitués au système soient tentés de répondre « Rejet » (k+1) ou « Toutes » (k+2). *Rejet* parce qu'en l'absence de certaines données, aucune solution n'apparaît comme la meilleure. *Toutes* parce qu'en l'absence de certaines données, toutes les solutions pourraient bien être correctes. Pour éviter cette confusion, le professeur expliquera ces différences de réponses à partir de plusieurs exemples. De plus, on introduira une *règle de priorité* entre ces diverses réponses possibles : k+3 l'emporte sur k+2, qui l'emporte sur k+1.

4. La solution générale k+4 : Absurdité(s) dans l'amorce

Cette consigne reprend les consignes k+1, k+2, k+3 et précise, *en outre*, que :

- Q = 1 : Une seule solution proposée est correcte
 G = 4 : Répondez par k+1 si aucune solution proposée n'est correcte (rejet)
 Répondez par k+2 si toutes les solutions proposées sont correctes
 Répondez par k+3 si les données de l'énoncé sont insuffisantes
 Répondez par k+4 si l'énoncé contient une absurdité
 R = 1 : Fournissez une seule réponse (ou omettez)

Exemple
 Question e

La calotte glaciaire de la planète Mercure se situe :

1. Au pôle Sud de cette planète.
2. Au pôle Nord de cette planète.

RC = 6

Répondre 3 (k+1) signifie : « La calotte glaciaire de la planète Mercure se situe ailleurs qu'en 1 ou 2. »

Répondre 6 signifie : « Cette question n'a pas de sens parce que son énoncé contient une erreur ». C'est la réponse correcte, car Mercure est trop près du soleil pour que puisse s'y développer une calotte de glace.

En l'absence de la consigne k+4, on est obligé de faire figurer, parmi les solutions proposées, une solution dénonçant l'absurdité de la question, comme dans les deux cas suivants (proposés par Sacré).

- Q = 1 : Une seule solution proposée est correcte
 G = 0
 R = 1 : Fournissez une seule réponse (ou omettez)

Question f

Le volume émergeant d'un iceberg est de 100 m³.
 Quel est son volume total et son poids sachant que le poids spécifique de la glace vaut 0,9 g/cm³ ?

1. Volume total = 900 m³; poids = 810 tonnes.
2. Volume total = 1000 m³; poids = 900 tonnes.
3. Volume total = 1000 m³; poids = 1000 tonnes.
4. Volume total = 900 m³; poids = 900 tonnes.
5. Ce problème est absurde, car un objet aussi lourd ne peut flotter sur la mer.

RC = 2

Q = 1 : Une seule solution proposée est correcte

G = 0

R = 1 : Fournissez une seule réponse (ou omettez)

Question g

Pourquoi un navire s'enfoncé-t-il davantage lorsque, venant de la mer, il pénètre dans un fleuve ?

1. Parce que l'eau d'un fleuve a un plus grand poids volumique que l'eau de mer.
2. Parce que l'eau d'un fleuve a un plus petit poids volumique que l'eau de mer.
3. Parce que l'eau de la mer est plus profonde que le fleuve.
4. La question contient une erreur, car le bateau ne s'enfoncé pas plus dans le fleuve.

RC = 2

La dernière solution ne manque pas de mettre la puce à l'oreille de l'étudiant, même peu vigilant. Il est simple d'éviter cet inconvénient grâce à la solution générale k+4.

Avantages de cette consigne

La solution générale k+4 permet de créer des questions à l'intersection des niveaux taxonomiques de l'analyse et de l'évaluation.

Ses dangers

Il arrive que le professeur construise des QCM dont la solution correcte est k+4, mais que l'absurdité figurant dans l'amorce passe inaperçue parce qu'il s'agit d'un détail ou, plus simplement, parce que l'étudiant ne connaît pas bien le problème. La question h offre un exemple de ce genre d'erreur de construction.

Q = 1 : Une seule solution proposée est correcte

G = 4 : Répondez par k+1 si aucune solution proposée n'est correcte

Répondez par k+2 si toutes les solutions proposées sont correctes

Répondez par k+3 si les données de l'énoncé sont insuffisantes

Répondez par k+4 si l'énoncé contient une absurdité

R = 1 : Fournissez une seule réponse (ou omettez)

Question h

Pouvez-vous signer en Belgique un chèque bancaire libellé en shillings autrichiens ?

1. Oui.
2. Non.

Pour le professeur, il s'agira de k+4 (absurdité dans l'amorce), parce que les shillings sont anglais et les schillings autrichiens. Les rares élèves qui remarquent cette erreur l'attribuent à une faute de dactylographie. On ne peut évidemment pas leur donner entièrement tort.

D. Discussion sur le principe des solutions générales

1. Une règle fondamentale

Lorsque plusieurs solutions générales sont adoptées dans une épreuve, une hiérarchie doit être établie entre elles.

Nous suggérons la hiérarchie $4 > 3 > 2 > 1$.

Ainsi, si k+4 et k+2 sont toutes deux correctes, c'est k+4 qui l'emporte. La règle serait qu'en cas de doute, k+4 prenne le pas sur toutes les autres, k+3 sur les deux qui restent et k+2 sur k+1.

2. Un exemple d'utilisation systématique

Voici des exemples extraits d'une épreuve de statistique descriptive que nous avons proposé en 1975 à des étudiants de l'Enseignement supérieur.

Les consignes k+1, k+2, k+3 et k+4 étaient simultanément adoptées. L'épreuve comportait quarante QCM, et le nombre de solutions proposées (k) pour chaque question variait entre 3 et 5. La moitié de ces questions avaient pour solution correcte une solution générale. Ces vingt questions furent moins bien réussies que les vingt autres, ce qui est normal puisqu'elles sollicitent davantage de vigilance et font appel à d'autres types de raisonnement, à notre avis plus élaborés.

Pour réussir certaines questions, une compréhension plus profonde des notions étudiées s'avérait nécessaire, comme le montrent les quelques exemples ci-dessous.

Q = 1 : Une seule solution proposée est correcte

G = 4 : Répondez par k+1 si aucune proposée n'est correcte

Répondez par k+2 si toutes les solutions proposées sont correctes

Répondez par k+3 si les données de l'énoncé sont insuffisantes

Répondez par k+4 si l'énoncé contient une absurdité

R = 1 : Fournissez une seule réponse (ou omettez)

Exemple

Question i

Un étudiant obtient un score de 12, alors que la moyenne des scores de sa classe vaut 9. Quel est son score z ?

1. +3
2. -3
3. +6
4. -6

La réponse correcte est 7, car on ne peut calculer z sans connaître l'écart type (manque de données : $k+3$).

Voici deux autres exemples.

(Voir consigne précédente)

Un étudiant obtient un score $z = +2$ dans une classe où la moyenne des scores vaut 14 et l'écart type vaut -4. Quel est son score brut ?

1. 16
2. 22
3. 12
4. 6

La réponse correcte est 8, car un écart type ne peut être négatif (absurdité : $k+4$)

(Voir consigne précédente)

Un étudiant obtient un score $z = -3$ dans une classe où la moyenne des scores vaut -5 et l'écart type vaut 1. Quel est son score ?

1. -4
2. -2
3. -8
4. -6

La réponse correcte est 3, bien que la tentation soit grande de répondre 8 sous le prétexte (erroné) qu'une moyenne ne peut être négative.

Dans l'épreuve qui vient d'être présentée, l'étudiant n'avait qu'une chance sur 7, 8 ou 9 (au moins) de pouvoir fournir la réponse correcte au seul hasard, au lieu d'une chance sur 3, 4 ou 5, ce qui est souvent le cas dans les questions n'utilisant pas systématiquement les solutions générales. Cette amélioration est valable pour toutes les QCM, que leur solution correcte soit une des solutions proposées ou une solution générale.

3. Critique des solutions générales

La difficulté de compréhension et d'utilisation des solutions $k+3$ et $k+4$ a déjà été signalée. Des remèdes ont été proposés.

Un deuxième problème tient aux difficultés que pourraient rencontrer les étudiants dans le codage de leurs réponses, c'est-à-dire dans le calcul de $k+g$.

On peut, en effet, craindre (mais la pratique ne le confirme pas) que les étudiants ne se trompent dans le calcul $k+1$, $k+2$, etc. D'aucuns ont proposé de consacrer le chiffre 9 au rejet de toutes les solutions, le chiffre 8 à la réponse « Toutes les solutions », le chiffre 7 à la réponse « Données manquantes », le chiffre 6 à la réponse « absurde ». Cette façon de faire pose autant de problèmes qu'elle n'en résout ; elle supprime, par exemple, la possibilité de créer des questions à six solutions ou plus, quand la consigne $k+4$ est en vigueur.

Certes, si $k = 6$, dans la procédure $(k+g)$ que nous proposons, la solution générale $k+4$ peut être exclue comme réponse correcte lorsque l'on n'admet que des réponses d'un chiffre. Cette procédure $k+g$ présente, par contre, des avantages d'ordre typographique et mécanographique (impression sur listing des analyses des questions).

Les deux systèmes sont donc relativement équivalents. On préférera tantôt l'un, tantôt l'autre, pour des raisons de commodité. On peut aussi adopter les initiales R (rejet), T (toutes), I (insuffisant) ou M (manquant), A (absurdité).

Enfin, les solutions générales obligent l'étudiant à se concentrer, non seulement sur les solutions qu'il a sous les yeux, mais en plus sur les possibilités non rappelées. Certains considèrent cela comme un défaut, une complication inutile. Nous estimons qu'il s'agit plutôt d'une qualité, car l'étudiant doit faire preuve d'une plus grande vigilance et d'un esprit plus critique.

Cependant, de telles procédures *ne doivent être introduites que très progressivement* : présenter d'abord la seule consigne $k+1$, puis, après quelques épreuves, la consigne $k+2$ et ainsi de suite.

Si l'on souhaite introduire de telles modalités d'évaluation, la coordination entre les professeurs d'une même école est des plus souhaitable.

Comme nous l'avons dit en commençant, cette procédure n'a encore été utilisée que dans l'enseignement supérieur, où elle fonctionne sans problème. Il reste à voir si elle pourra éventuellement être appliquée, mutatis mutandis, dans l'enseignement secondaire, voire primaire en prenant certaines précautions. Les premières épreuves de ce type pourraient, par exemple, être présentées aux élèves sous forme de jeu.

4. Solutions générales et questions VRAI-FAUX

L'exemple ci-dessous montre quelles difficultés peuvent surgir si l'on applique les quatre solutions générales à une question VRAI-FAUX.

Question 1

Q = 1 : Une seule solution proposée est correcte

G = 4 : Répondez par $k+1$ si aucune solution proposée n'est correcte

Répondez par $k+2$ si toutes les solutions proposées sont correctes

Répondez par $k+3$ si les données de l'énoncé sont insuffisantes

Répondez par $k+4$ si l'énoncé contient une absurdité

R = 1 : Fournissez une seule réponse (ou omettez)

Le Zaïre est une ancienne colonie hollandaise.

1. Vrai.
2. Faux.

La réponse correcte est évidemment, 2 (Faux). Néanmoins, certains pourraient répondre 6 ($k+4$), en pensant – avec raison – que le corps de la question contient une absurdité. Or c'est toujours le cas pour les questions VRAI-FAUX dont la réponse correcte est « faux ».

Dès lors, *le recours aux solutions générales $k+1$, $k+2$ et $k+4$* est à éviter avec ces questions. Une exception de ce type peut être faite pour la consigne $k+3$ (trois réponses sont possibles) :

- 1 = vrai ;
- 2 = faux ;
- 3 = données insuffisantes pour répondre.

Cette consigne permet d'aborder certains contenus « indécidables » comme dans l'exemple suivant :

Question m (voir consigne précédente)

La vie n'existe que dans notre galaxie.

1. Vrai.
2. Faux.

5. D'autres solutions générales

Il est possible de créer d'autres solutions générales ($k+5$ etc.) et l'avenir en verra sans doute naître qui seront plus appropriées à certains domaines non envisagés ci-dessus.

Luttgens nous a par exemple suggéré :

« Ceci est une question d'opinion ou de goût », permettant de traiter les questions d'opinion, et « L'état actuel de nos connaissances ne permet pas de trancher », permettant de traiter les questions non décidables, ou encore « Cette question a des implications, ou des présupposés philosophiques, politiques, ou sociaux... que je ne partage pas ».

Pour cette dernière solution, Luttgens nous a donné quelques exemples de questions :

Q = 1 : Une seule solution proposée est correcte

G = 5 : Répondez par $k+1$ si aucune solution proposée n'est correcte

Répondez par $k+2$ si toutes les solutions proposées sont correctes

Répondez par $k+3$ si les données de l'énoncé sont insuffisantes

Répondez par $k+4$ si l'énoncé contient une absurdité

Répondez par $k+5$ si vous ne partagez pas les valeurs énoncées.

R = 1 : Fournissez une seule réponse (ou omettez)

Question n – La vermine capitaliste a été éliminée de la République XYZ en 1.A
2.B
3.C

Question o – Après notre résurrection, quel aspect aura notre corps ? 1.A
2.B
3.C

Question p – La preuve définitive de la non existence de Dieu a été administrée par 1.A
2.B

Dans le même ordre d'idées, nous suggérons : « Cette question est inacceptable » (solution qui pourrait être utile dans le cours de morale).

Exemple : (QGR = 151) :

Question q – Avec quel instrument faut-il torturer quelqu'un pour que ...

Question r – Jusqu'où faut-il affamer un peuple pour que ...

Question s – Combien de morts pour l'exemple doit-on faire quand ...

On pourrait continuer ainsi avec des solutions du type : « Cette question est indiscreète ... » etc.

L'intérêt d'utiliser de telles solutions générales est, comme on l'a déjà expliqué au chapitre 1 à propos des sollicitations implicites, d'habituer l'élève à certains types de réponses et d'entretenir sa vigilance intellectuelle ou morale.

*
* *

Nous venons de suggérer des moyens de pré-coder (avec des chiffres) les solutions générales. Dans le cas d'un testing par ordinateur, on peut permettre aux étudiants de s'exprimer (« aucune n'est correcte » ou « tout est faux » ou « c'est inadmissible ... » dans ses propres termes. Le programme doit alors « interpréter » les objections (prévues ou non prévues par le programmeur) faites par l'étudiant.

On le voit, un pas supplémentaire est ainsi fait dans l'encouragement des réponses spontanées qui, comme nous le signalons au chapitre 1, sont le but à atteindre.

E. Exemple d'épreuves aux solutions générales

A titre d'exemple, cinq questions de géométrie, cinq questions de géographie et cinq questions relatives à l'ordre alphabétique sont présentées ci-après. Le lecteur est invité à y répondre dans la grille, puis à comparer ses réponses aux solutions correctes qui suivent.

	Géométrie					Géographie					Alphabet				
QUESTIONS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
REPONSES															

Consigne

Dans les QCM qui suivent, il faut répondre par :
 k+1 pour « Aucune de ces solutions n'est correcte (Rejet) »,
 k+2 pour « Toutes ces solutions sont correctes »,
 k+3 pour « Il est impossible de répondre parce que au moins une donnée manque dans l'énoncé. »,
 k+4 pour « Une absurdité dans l'énoncé rend toute la question sans objet ».
 Attention ! - La réponse k+4 a priorité sur les trois autres.
 - Ces solutions générales, sauf la solution k+3, ne sont pas valables pour les questions VRAI-FALUX. Pour ce type de questions, les seules réponses possibles sont donc 1, 2 et 5.

1. Géométrie

Question 1 : Un triangle isocèle a une base de 10 cm. Quel est son périmètre ?

1. 28 cm 2. 30 cm 3. 32 cm.

Question 2 : Un triangle équilatéral a un côté de 10 cm. Quel est son périmètre ?

1. 28 cm 2. 30 cm 3. 32 cm.

Question 3 : Un triangle isocèle a deux côtés de 5 m et un côté de 6 m. Quelle est sa surface (en m²) ?

1. 8 m² 2. 10 m² 3. 15 m² 4. 30 m²

Question 4 : Dans un triangle équilatéral, quelle droite part de chaque sommet pour aboutir au milieu du côté opposé ?

1. La hauteur 2. La médiatrice 3. La bissectrice.

Question 5 : Les trois côtés d'un triangle mesurent respectivement 6 m, 2 m et 3 m. Quel est son périmètre ?

1. 11 m 2. 12 m 3. 15 m 4. 18 m

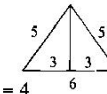
Clé de correction

QUESTIONS	1	2	3	4	5
REPONSES CORRECTES	6	2	5	5	8
OU k +	3	-	1	2	4

Commentaires sur certaines questions

Question 1 : Il est impossible de répondre, puisqu'on ne connaît pas la longueur des autres côtés (R. C = k+3).

Question 3 : La réponse correcte est 12 m²
 En effet,



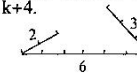
$$\text{or } x^2 + 3^2 = 5^2$$

$$x^2 = 5^2 - 3^2 = 25 - 9 = 16 \text{ et } x = 4$$

R.C. = k+1 puisque 12 m² ne figure pas parmi les solutions proposées.

Question 4 : Les trois solutions sont correctes. Il faut répondre par k+2.

Question 5 : Il est impossible d'obtenir un triangle à partir de ces trois côtés. Il faut donc répondre par k+4.



2. Géographie

Question 6 : Laquelle des situations (en latitude et en longitude) est correctement indiquée ?

- Point A : 80° de latitude Nord et 190° de longitude Est.
- Point B : 120° de latitude Sud et 120° de longitude Ouest.
- Point C : 30° de latitude Est et 120° de longitude Ouest.
- Point D : 50° de latitude Nord et 90° de longitude Sud.
- Point E : -10° de latitude Sud et 30° de longitude Est.

Question 7 : De quel endroit de la terre peut-on dire qu'il n'a pas de longitude et se situe à une latitude de 90°.

- Le pôle Nord (géographique).
- Le pôle Sud (géographique).

Question 8 : Deux villes situées sur le même méridien sont distantes de 50° de latitude. Quelle ville connaît le climat le plus froid ?

- Celle du Nord
- Celle du Sud

Question 9 : Pour calculer la longueur approximative d'un degré de longitude à la latitude de Paris.

1. On divise 40 000 km par 90.
2. On divise 40 000 km par 180.
3. On divise 40 000 km par 360.

Question 10 : Deux villes sont situées sur le même parallèle. Entre la ville A, la plus proche de l'équateur, et la ville B, la plus éloignée de l'équateur, il y a 100° de latitude. Quelle ville connaît le climat le plus chaud ?

1. A.
2. B.

Clé de correction

QUESTIONS	6	7	8	9	10
REPONSES CORRECTES	6	4	5	4	6
OU k +	1	2	3	1	4

Commentaires sur certaines questions

Question 8 : C'est impossible à dire, car il se pourrait qu'elles soient de part et d'autre de l'équateur, toutes deux dans l'hémisphère Nord, ou toutes deux dans l'hémisphère Sud. Bref, il manque des données pour pouvoir répondre. La réponse correcte est k+3.

Question 9 : Aucune des quatre solutions n'est correcte, car un degré de longitude représente, à la latitude de Paris, une distance bien moins grande qu'à l'équateur, où la formule 3 serait valable. La réponse correcte est k+1.

Question 10 : Si ces deux villes se situent sur le même parallèle, elles ne peuvent être distantes de 100° de latitude ; l'une ne peut être plus proche de l'équateur que l'autre. Il existe donc une absurdité (au moins) dans l'énoncé. La réponse correcte est k+4.

3. Alphabet

Un annuaire officiel des téléphones contient plusieurs listes alphabétiques (une par zone, une par rubrique des pages jaunes, etc.).

Question 11 : Voici un extrait d'une liste. On cherche le nom Martius. On doit chercher :

1. Avant cet extrait.
2. Après cet extrait.

a

21 00 50 Martius J. (111)
21 00 51 Martius J. (112)
21 00 52 Martius J. (113)
21 00 53 Martius J. (114)
21 00 54 Martius J. (115)
21 00 55 Martius J. (116)
21 00 56 Martius J. (117)
21 00 57 Martius J. (118)
21 00 58 Martius J. (119)
21 00 59 Martius J. (120)

Question 12 : Voici un extrait d'une liste. Le nom Bameray se trouve-t-il dans cette liste ?

1. Oui.
2. Non.

b

22 00 15 Bameray J. (15)
22 00 16 Bameray J. (16)
22 00 17 Bameray J. (17)
22 00 18 Bameray J. (18)
22 00 19 Bameray J. (19)
22 00 20 Bameray J. (20)
22 00 21 Bameray J. (21)
22 00 22 Bameray J. (22)
22 00 23 Bameray J. (23)
22 00 24 Bameray J. (24)
22 00 25 Bameray J. (25)
22 00 26 Bameray J. (26)
22 00 27 Bameray J. (27)
22 00 28 Bameray J. (28)
22 00 29 Bameray J. (29)
22 00 30 Bameray J. (30)

Question 13 : Voici un extrait d'une liste. Le nom Leriche se trouve-t-il dans cette liste ?

1. Oui.
2. Non.

c

23 00 15 Leriche J. (15)
23 00 16 Leriche J. (16)
23 00 17 Leriche J. (17)
23 00 18 Leriche J. (18)
23 00 19 Leriche J. (19)
23 00 20 Leriche J. (20)
23 00 21 Leriche J. (21)
23 00 22 Leriche J. (22)
23 00 23 Leriche J. (23)
23 00 24 Leriche J. (24)
23 00 25 Leriche J. (25)
23 00 26 Leriche J. (26)
23 00 27 Leriche J. (27)
23 00 28 Leriche J. (28)
23 00 29 Leriche J. (29)
23 00 30 Leriche J. (30)

Question 14 : Voici un extrait d'une liste. Le nom Honius se trouve-t-il dans cette liste ?

1. Oui.
2. Non.

d

24 00 15 Honius J. (15)
24 00 16 Honius J. (16)
24 00 17 Honius J. (17)
24 00 18 Honius J. (18)
24 00 19 Honius J. (19)
24 00 20 Honius J. (20)
24 00 21 Honius J. (21)
24 00 22 Honius J. (22)
24 00 23 Honius J. (23)
24 00 24 Honius J. (24)
24 00 25 Honius J. (25)
24 00 26 Honius J. (26)
24 00 27 Honius J. (27)
24 00 28 Honius J. (28)
24 00 29 Honius J. (29)
24 00 30 Honius J. (30)

Question 15 : Voici un extrait d'une liste. Si l'on devait ajouter le nom Girioli, il figurerait :

1. Sur cette liste.
2. Sur cet extrait.
3. Entre Giriots et Girin.

e

25 00 15 Girioli J. (15)
25 00 16 Girioli J. (16)
25 00 17 Girioli J. (17)
25 00 18 Girioli J. (18)
25 00 19 Girioli J. (19)
25 00 20 Girioli J. (20)
25 00 21 Girioli J. (21)
25 00 22 Girioli J. (22)
25 00 23 Girioli J. (23)
25 00 24 Girioli J. (24)
25 00 25 Girioli J. (25)
25 00 26 Girioli J. (26)
25 00 27 Girioli J. (27)
25 00 28 Girioli J. (28)
25 00 29 Girioli J. (29)
25 00 30 Girioli J. (30)

Clé de correction

QUESTIONS	11	12	13	14	15
REPONSES CORRECTES	3	6	5	2	5
OU k +	1	4	3	-	2

Commentaires sur les questions portant sur l'ALPHABET

Question 11 : Il n'est pas nécessaire d'examiner d'autres extraits (ni avant, ni après) pour se rendre compte que Martius n'est pas dans la liste. Il se trouverait entre Martiny et Martucci. Aucune des deux solutions n'est donc correcte. La réponse correcte est k+1.

Question 12 : Le prétendu extrait de l'annuaire téléphonique est incohérent : Beauvuin précède Bataille, Baugnet précède Banse. Ou cet annuaire a été mal construit, ou il s'agit d'un faux extrait (ce qui est le cas). L'énoncé contient donc une absurdité. La réponse correcte est k+4.

Question 13 : Il est impossible de répondre à la question sur la base de ce seul extrait : il manque des informations. La réponse correcte est k+3.

Question 14 : L'extrait permet de répondre : Honius ne se trouve pas dans cet annuaire, car il serait entre Honin et Honnay. La réponse correcte est 2.

Question 15 : Les trois solutions proposées sont correctes. La réponse correcte est k+2.

F. Les codes de chaque QCM

On pourrait s'étonner de l'attention que nous avons consacrée à des problèmes de codage. Il sont utiles dans la perspective de banques de questions gérées par ordinateur, pour trois raisons citées ci-après. Mais, avant tout, il faut se garder de confondre les codes (QGRL, etc.) de la consigne et ceux de la question.

En effet, la consigne peut annoncer que des solutions générales sont en vigueur, mais la solution correcte peut être l'une des solutions proposées. De même, la consigne peut annoncer combien de solutions proposées sont correctes, mais ne précise évidemment pas lesquelles. Plusieurs informations sur la QCM sont utiles, voire nécessaires.

1. Nécessité pour la correction.

Le correcteur d'un test procédant par QCM peut corriger les réponses sans voir les QCM, à condition de disposer du numéro de la solution correcte de chaque QCM... ou des solutions correctes le cas échéant. Il doit

parfois disposer de plus d'informations encore : le nombre de solutions proposées (k) (quand il s'interroge sur les chances de fournir la solution correcte au hasard), ou le nombre de questions avec lesquelles chaque QCM est solidaire (quand on lie l'attribution de points à la réussite de plusieurs questions).

2. Nécessité pour la construction de test.

Nous avons vu que la plupart des QCM pourraient être utilisées avec diverses consignes, mais pas toutes ! Imaginons que nous disposions d'une banque de QCM et que nous voulions construire un test utilisant une consigne unique. Si les QCM ont été adéquatement codées, on doit pouvoir demander à un ordinateur de désigner toutes celles qui sont compatibles avec une consigne donnée.

La situation inverse peut se produire : on peut avoir choisi des QCM pour figurer dans un test, puis se demander quelles(s) consigne(s) sont possibles pour l'ensemble de ces questions.

3. Utilité pour l'interprétation des résultats.

Il peut être intéressant de vérifier si les QCM qui présentent la solution correcte (PSC) sont mieux réussies que celles qui la cachent (par exemple celles où la solution correcte est « aucune »). Un des codes de la question devrait permettre de différencier ces QCM des autres et d'établir les comparaisons automatiquement (par ordinateur).

4. Codes proposés.

K = le nombre de solutions proposées (figurant à la suite de l'annonce, donc sans compter d'éventuelles solutions générales).

NSC = le nombre de solutions correctes. Ce nombre peut être annoncé (consignes Q=1 et Q=2) ou non (consignes Q=3 et Q=4).

NQL = le nombre de questions liées dont fait partie la QCM. Rappelons que le code L de la consigne indique uniquement la position (première, dernière, intermédiaire) de la QCM dans une série.

PSC = Présentation (PSC=1) ou non (PSC=0) de la solution correcte.

Soit RC = le numéro de la (ou les) réponse(s) correcte(s).

Soit VR = les valeurs des k+g réponses.

Si des solutions générales sont en vigueur, RC peut être supérieur à k.

Si le nombre de solutions correctes (NSC) est supérieur à 1, alors on doit fournir non pas *le* numéro de la réponse correcte, mais la liste des valeurs des k solutions proposées. Rappelons que les solutions générales impliquent que l'on fournisse une seule réponse (code R=1).

Exemple : Consigne QR=11

Quelle est la capitale de l'Italie ?

1. Venise.
2. Milan.
3. Naples.
4. Florence.
5. Autre.

Codes de la question (K, NSC, NQL, PSC, RC) = 5, 1, 0, 0, 5

Exemple : Consigne QR=31

Q3 = une ou plusieurs solutions proposées sont correctes.

R1 = fournissez *une* réponse.

Choisissez une ville traversée par le Danube.

1. Berlin.
2. Prague.
3. Vienne.
4. Budapest.
5. Varsovie.
6. Belgrade.

Codes de la question (K, NSC, NQL, PSC, RC) = 6, 3, 0, 1, 0, 0, 1, 1, 0, 1
} les RC

Conclusions

Le tour d'horizon qui vient d'être fait ne concerne que la *conception* des QCM.

Tout naturellement, ces préoccupations devraient être prolongées par une réflexion sur leur *utilisation*, et l'*interprétation* des données qu'elles permettent de recueillir.

C'est pourquoi nous avons préparé un second ouvrage traitant des sujets suivants :

- *La fidélité des épreuves scolaires*, et la contribution des QCM à l'amélioration de cette fidélité. Sont développés les formules de calcul de la fidélité, le problème du nombre optimal de questions dans un test ou de solutions dans une QCM, la fraude et sa prévention, et enfin l'entraînement des étudiants à répondre à des QCM.
- *La notion de difficulté d'une question*. Il s'agit de distinguer trois approches du concept : théorique, subjective et expérimentale. Le rôle du hasard doit être analysé, ainsi que l'influence du contexte et de la compétence de l'étudiant sur la facilité.
- *Les indices du pouvoir discriminatif d'une question*. Quels sont leurs principes et la façon de les interpréter ? Comment les calculer (avec et sans ordinateur) ? Sont-ils comparables les uns aux autres ?
- *Les formules de notation (scoring)*. On connaît la fameuse « correction for guessing ». Quel est son principe fondamental ? Qu'implique-t-elle sur les plans théorique et pratique ? Sous quelles formes différentes existe-t-elle ? Faut-il y recourir et pourquoi ?

Nous espérons que ces deux ouvrages mettent des réponses précises à la disposition des enseignants qui veulent se doter d'outils d'évaluation de qualité.

Bibliographie

- BENTLER, P.M., JACKSON, D.N., and MESSICK, S., 1971, Identification of content and Style : A two dimensional interpretation of acquiescence. *Psychological Bulletin*, 76, 186-204.
- BERGLUND, G.W. (1969), Effect of knowledge of results on retention, *Psychol. Sch.*, 6, 420-427.
- BETZ, N.E. & WEISS, D.J. (1976 (a)), Effects of immediate knowledge of results and adaptive testing on ability test performance. *Research Report, 76-3*, Psychometric Methods Program, University of Minnesota.
- BISHOP, A.J., KNAPP, T.R. & MacINTYRE, D.I. (1969), A comparison of the results of open-ended and multiple-choice versions of a mathematics test, *Int. Jour. Educ. Sci.*, 3, 147-154.
- BLOOM, B.S. et al. (1971), *Handbook on formative and summative evaluation of student learning*, Mc Graw Hill, New York.
- BOARD, C. & WHITNEY, D.R. (1972), The effect of selected poor itemwriting practices on test difficulty, reliability and validity, *Jour. Educ., Measmt.*, 9, 225-233.
- BOULANGER, R. (1968), Qu'est-ce que la cybernétique ? in Le dossier de la cybernétique, *Marabout Université*, 1968, 10-27.
- BONBOIR, A. (1960), L'arithmétique au niveau de la sixième primaire, Commission consultative de Pédagogie universitaire, Bruxelles.
- BRANDENBURG, D.C. & WHITNEY, D.R. (1972), Matched Pair True-False Scoring : Effect on Reliability and Validity, in *Journal of Educ. Measur.*, vol. 9, N° 4.
- CAVERNI, J.P., FABRE, J.M. et NOIZET, G. (1975), Dépendance des évaluations scolaires par rapport à des évaluations antérieures : études en situation simulée, *Le travail humain*, 38, 213-222.
- CHASE, C. (1964), Relative length of option and response set in multiple choice items, *Educ. and Psychol. Meas.*, vol. 24, 1, 861-866.
- CHOPPIN, B. (1970), *An IEA Study of Guessing. A Proposal*, Stockholm, International Association for the Evaluation of Educational Achievement. Unpublished Memorandum, IEA/TR/9.
- CHOPPIN, B.H. (1974 (a)), *The Correction for Guessing on Objective Tests*, IEA Monograph Studies, n° 4, Stockholm.
- CLARKE (Ed.) (1975), *Chemistry : Objective and Completion Tests for O Level*, J. MURRAY Publisher Ltd.
- COLLET, L.S., 1971, Elimination scoring : an empirical evaluation, *Journ. Educ. Meas.*, 8, 209-214.
- COOMBS, C.H., MILHOLLAND, J.E. & WOMER, F.B. (1956), The assessment of partial knowledge, *Educ. Psychol. Measmt.*, 16, 13-37.
- COPELAND, D.A. (1972), Should chemistry students change answers on multiple-choice tests ? *Jour. Chem. Educ.*, 49, 258.
- COUGH, A. and KENISTON K., 1960, Yea sayers and nay sayers : Agreeing response set as a personality variable. *Journal of Abnormal Social Psychology*, 60, 151-174.
- CREHAN, K.D. (1974), Item analysis for teacher-made mastery tests, *Jour. Educ. Measmt.*, 11, 255-262.
- CRONBACH, L.J., 1942, Studies of acquiescence as a factor in the true-false test, *Journal of Educational Psychology*, 33, 401-415.
- DE BAL, R., DE LANDSHEERE, G., PAQUAY-BECKERS, J. (1976), *Construire des échelles descriptives d'évaluation*, Coll. « Recherche en éducation », Bruxelles, Direction générale de l'Organisation des Etudes du Ministère de l'Education nationale.

- DEBOT, F. et LECLERCQ, D. (1978), *Guide pratique du STEP* (Système de Traitement automatique d'Évaluations Pédagogiques), document roncotypé, Université de Liège.
- DE LANDSHEERE, G. (1973), *Le test de closure, Mesure de la lisibilité et de la compréhension*, Paris, F. Nathan; Bruxelles, Labor.
- DE LANDSHEERE, G. (1979), *Dictionnaire de l'évaluation et de la recherche en éducation*, Paris, PUF.
- DE LANDSHEERE, G. et V. (1982), *Définir les objectifs de l'éducation*, Paris, PUF, Liège, G. Thone, 5e éd.
- DE LANDSHEERE, G. (1980), *Évaluation continue et examens. Précis de docimologie*, Paris, F. Nathan; Bruxelles, Labor 5e éd.
- DE LANDSHEERE, G. (1982), *Introduction à la recherche en éducation*, Paris, A. Colin, Liège, G. Thone, 6e éd.
- D'HAHNAUT, L. (1973), Étude d'une nouvelle variable pour l'analyse statistique des expériences pédagogiques, *Bulletin de Psychologie*, 305, vol. XXVI, p. 622-630.
- DIAMOND, J.J. & EVANS, W.J. (1972), An investigation of the cognitive correlates of test-wisness, *Jour. Educ. Measmt.*, 9, 145-150.
- DRESSEL, P. & SCHMID, J. (1953), Some modifications of the multiple-choice items, *Educ. and Psychol. Measmt.*, vol. 13, pp. 574-595.
- DUDLEY, H.A.F. (1973), Multiple-choice tests, *Lancet.*, 2, 195.
- DUDYCHA, A.L. & CARPENTER, J.B. (1973), Effects of item format on item discrimination and difficulty, *Jour. Appl. Psychol.*, 58, 11-121.
- DUNN, T.F. & Goldstein, L.G. (1959), Test difficulty, validity and reliability as functions of selected multiple-choice item construction principles, *Educ. Psychol. Measmt.*, 19, 171-179.
- EBEL, R.L. (1965), *Measuring Educational Achievement*, Englewood Cliffs, N.J., Prentice Hall, A.
- EBEL, R.L. (1969), Expected reliability as a function of choices per item, *Educ. Psychol. Measmt.*, 29, 565-570.
- EBEL, R.L. (1970), The case for true-false test items, *School Rev.*, 78, 373-39.
- EBEL, R.L. (1971), How to write true-false test items, *Educ. Psychol. Measmt.*, 31, 417-426.
- EKLUND, H. (1968), Multiple Choice and Retention, Almqvist and Wiksells, Uppsala.
- FABRE, J.M. (1975), *Docimologie expérimentale et évaluation par questionnaires : étude du jugement multiple et de l'autopondération*, Thèse de doctorat de troisième cycle en psychologie, Université de Provence, roncotypé, 224 p.
- FABRE, J.M. et NOIZET, G. (1977a), Confiance attachée aux réponses à des questions à choix multiple, *Journal de Psychologie normale et pathologique*, 74, 335-362.
- FABRE, J.M. et NOIZET, G. (1977b), Jugement du vrai et du faux sur des énoncés affirmatifs et négatifs, in *Psychologie expérimentale et comparée*, Hommage à Paul Fraise, Paris, PUF, 411-426.
- FROM, J. (1970), Plaidoirie pour un enseignement meilleur ou introduction à l'enseignement intégral, in *MDT* n° 3, 11-12.
- GAGNE, R.M. (1965), *The conditions of Learning*, New York, Holt, Rinehard et Winston.
- GIBB, B.G., 1964, Test-wisness as secondary cue response (Doctoral Dissertation, Stanford University) Ann Arbor, Michigan : University Microfilms, n° 64-7643.
- GRÖNLUND, N., 1965, Measurement and evaluation in teaching, N. Y., Ed. Mc Millan.
- QUILFORD J.P. (1967), *The nature of human intelligence*, New York, Mc Graw Hill.
- HALLEUX-HENDRICK, J. (1969b), Construction des questions à choix multiple : une seule solution correcte ?, in *Revue Belge de Psychologie et de Pédagogie*, Tome XXXI, n° 127, 113-125, Bruxelles.

- HALLEUX-HENDRICK, J. (1969a), L'attitude de l'élève de l'enseignement secondaire face aux questions à choix multiple, in *Revue Belge de Psychologie et de Pédagogie*, Bruxelles, tome XXXI, n° 127, 87 et ss.
- HARDY, J.L., DIGHAYE, J., BERGER, Y., BRUXANT, M., DEWERCHIN, M., GILBERT, M., LEFEVRE, M., DENIS, B., LECLERCQ, D., 1982, Modules d'auto-évaluation en chimie, Ministère de l'Éducation Nationale, Organisation des Etudes.
- HARDY, J.L. (1979), *Approche expérimentale du comportement d'estimation et de sa mesure. Implications pour l'enseignement*, Mémoire de licence en Sciences de l'éducation, Université de Liège.
- HARDY, J.L. (1981), Using computer-based feedbacks to improve estimation ability, in Lewis and Tagg (Eds), *Computers in Education*, North Holland Publishing Company, IFIP, 317-323.
- HENRY, G., 1977, Comment mesurer la lisibilité, Bruxelles : Labor.
- HOUZIAUX, M.O. (1972), *Vers l'enseignement assisté*, Paris, PUF.
- HUGHES, H. & TRIMBLE, E. (1965), The use of complex alternatives in multiple-choice items, in *Educ. and Psychol. Measmt.*, 25, 1.
- KAESS, W. et ZEAMAN, D. (1960), Positive and negative knowledge of results on a Pressey-type punchboard, *Journal of Experimental Psychology*, 60, 12-17.
- KARRAKER, R.J. (1967), Knowledge of results and incorrect recall of plausible multiple-choice alternatives, *Jour. Educ. Psychol.*, 58, 11-14.
- LABORIT, H. (1968), La cybernétique et la machine humaine, in *Le dossier de la cybernétique. Marabout Université*, 10-27.
- LARKINS, A.G., and SHAVER, J.P., 1968, Comparison of yes-no matched-pairs, and all-no scoring of a first-grade economics achievement test. ERIC Report n° ED029701 - Utah State University, Logan, Utah.
- LECLERCQ, D. (1971), Une banque de questions pour l'enseignement, *Education*, n° 132.
- LECLERCQ, D. (1976), La fonction régulatrice de l'évaluation vue sous l'angle de l'implication de l'étudiant, in *Education*, n° 159.
- LECLERCQ, D. (1978), Un module d'auto-évaluation ou Comment impliquer l'étudiant dans la régulation de ses apprentissages, in *Education*, n° 165, pp. 59-73.
- LECLERCQ, D., DONNAY, J., DE BAL, R. (1977), Construire un cours programmé, 2e édition, Labor Bruxelles, Nathan Paris.
- LECLERCQ, D. (1978), La question à choix multiple, Un outil d'hier ou de demain ? *Les Cahiers Pédagogiques*, IPN, Paris, n° 162, 3-6.
- LECLERCQ, D., 1969, PACELBRO : Une technique originale d'enseignement programmé, in *Education*, Bruxelles, 1969, 118, 3-8.
- LECLERCQ, D., 1970, Étude comparative du rendement d'un cours programmé portant sur des exercices correctifs, in *Education*, 122, 3-7.
- LECLERCQ, D., 1982, Les séquences d'entraînement à la Résolution de Problèmes, (SERP), L.P.E. Université de Liège.
- LECLERCQ, D., REYMEY, J.M., MAHAUX, E., BOVY A., DE CNOP C., DOHOOGNE, J., RUHWIEDEL, M., SURKEYN, J., MARBAISE, C., 1982, Modules d'auto-évaluation en physique, Bruxelles : Ministère de l'Éducation Nationale, Organisation des Etudes.
- LECLERCQ, D., SIMON, J.P., BOVY, A., CHARNEUX, P., DE BIES, N., ETIENNE, C., HUMBLET, J., MONFILS, G., 1982, Modules d'auto-évaluation en biologie, Bruxelles : Ministère de l'Éducation Nationale, Organisation des Etudes.
- MASSENGILL, H. & SHUFORD, E. (1967), What pupils and teachers should know about guessing, Technical Report SMC R-7, Lexington (Mass).

- MARSHALL, J. and HALES, L. (1971), Classroom test construction, Addison-Wesley.
- MARSHALL, J. and HALES, L. (1972), Essentials of testing, Addison-Wesley.
- MAZER, A. (1975), Les QCM dans les examens de médecine, Editeur La Sirmar, Tours (France).
- McMORRIS, R.F., BROWN, J.A., BROWN, G.W. & PRUZEK, R.M. (1972), Effects of violating item construction principles, *Jour. Educ. Measmt.*, 9, 287-296.
- MICHAEL, J.J. (1968), The reliability of a multiple-choice examination under various test-taking instructions, *Journ. of Educ. Measmt.*, 5, 307-314.
- MIKLICH, D.R., and GORDON G., 1968, Test-taking carefulness vs. acquiescence response set on true-false examinations, *Educ. and Psychol. Measurement*, 28, 545-548.
- MILLMAN, J., BISHOP, C.H. and EBEL R., 1965, An analysis of test-wisness. *Educational and Psychological Measurement*, 25, 707-726.
- NILSSON, I. & WEDMAN, I. (1978), On test-wisness and some related constructs, *Scand. Jour. Educ. Res.*, 20, 25-40.
- NOIZET, G. et CAVERNI, J.P. (1978), Psychologie de l'évaluation scolaire, Paris, PUF.
- NOIZET, G. et FABRE, J.M. (1975), Etude docimologique des questionnaires à choix multiple (QCM) : Perspectives de recherche, *Scientia Paedagogica Experimentalis*, vol. 12, 38-62.
- OWENS, R., HANNA, G., COPPEDGE, F. (1970), Comparison of multiple-choice tests using different types of distractor selection techniques, *Journal of Educ. Measmt.*, vol. 72.
- PAYNE, R.W. & PENNYCUICK, D.B. (1975), *Multiple Choice Questions on Advanced Level Mathematics*, Bell, London.
- PETERSON, C.C. & PETERSON, J.L. (1976), Linguistic determinants of the difficulty of true-false test items, *Educ. Psychol. Measmt.*, 36, 161-164.
- PIERON, H. (1969), *Examens et docimologie*, Paris, PUF.
- PIKLER E. (1971), *Se mouvoir en liberté dès le premier âge*, Paris, PUF.
- PIPPERT, R. (1966), Final note on the changed answer myth, *Clearing House*, 38, 165-166.
- POWEL, J.C. & ISBISTER, A.G. (1974), A comparison between right and wrong answers on a multiple choice test, *Educ. Psychol. Measmt.*, 34, 499-509.
- POWEL, J.C., 1968, The interpretation of wrong answers from a multiple choice test, *Educ. and Psychol. Measur.*, 28, 403-412.
- PRESSEY, S. (1926), A simple apparatus which gives tests and scores and teaches, *School and Society*, vol. 23, n° 586, reprod. in R. GLASER, *Teaching Machines and programmed learning*, Nat. Educ. Assoc., Washington, 1961, 3rd Ed., pp. 35-40.
- PRESTON, R.C. (1965), Multiple-choice test as an instrument in perpetuating false concepts, *Educ. Psychol. Measmt.*, 25, 111-116.
- PYRCZAK, F. (1972), Objective evaluation of the quality of multiple-choice test items designed to measure comprehension of reading passages, *Read. Res. Quart.*, 8, 62-71.
- PYRCZAK, F. (1974), Passage-dependance of items designed to measure the ability to identify the main ideas of paragraphs : Implications for validity, *Educ. Psychol. Measmt.*, 34, 343-348.
- QUINN, B. (1975), A technical report on the multiple-choice tests set by the London GCE Board 1973 and 1974, University of London School Examinations Department, London, Unpublished manuscript.
- RABINOWITZ, F.M. (1970), Characteristic sequential dependencies in multiple-choice situations, *Psychol. Bull.*, 74, 141-148.
- RAMOS, R.A. & STERN, J. (1973), Item behavior associated with changes in the number of alternatives in multiple-choice items, *Jour. Educ. Measmt.*, 10, 305-310.

- REMMERS, H. & GAGE, N. (1955), *Educational Measurement and Evaluation*, New York, Harper.
- ROLLER, S. (1967), Le test de dictionnaire CL4, IRDP, Neuchâtel.
- ROWLEY, G.L. (1974), Which examinees are most favoured by the use of multiple-choice tests? *Jour. Educ. Measmt.*, 11, 15-23.
- SCRIVEN, M. (1967), *The Methodology of Evaluation* in R. STAKE (Ed.), *Curriculum Evaluation*, Monograph Series on Evaluation, n° 1, AERA, Rand Mc Nally, Chicago.
- SKINNER, B.F. (1971), *Analyse expérimentale du comportement* (trad. M. RICHELLE), Paris, Dessart.
- SKINNER, B.F., 1958, Reinforcement Today, *American Psychologist*, 13 (3).
- SKINNER, B.F., 1961, Why we need teaching machines, *Harvard Educational Review*, 31, 377-398.
- SKINNER, B.F., 1968, *La révolution scientifique de l'enseignement*, Bruxelles : Dessart.
- SLAKTER, M., KOELHER, R. & HAMPTON, S. (1970), Grade level, sex and selected aspects of test-wisness, *Journ. of Educ. Measmt.*, vol. 7, 119-ss.
- SLAKTER, M., KOELHER, R. & HAMPTON, S. (1970), Learning test-wisness by programmed texts, *Journ. of Educ. Measmt.*, vol. 7, 247-254.
- SMITH, R. (1970), An empirical investigation of complexity and Process in multiple-choice items, in *Journ. of Educ. Measmt.*, vol. 7, 33-41.
- STRANG, H.R. & RUST, J.O. (1973), The effects of immediate knowledge of results and task definition on multiple-choice answering, *Jour. Exper. Educ.*, 42, 77-80.
- STRANG, H.R. (1977), The effect of technical and unfamiliar options on guessing on multiple-choice test items, *Journ. of Educ. Measmt.*, vol. 14, 253-259.
- TERRANOVA, C. (1969), The effects of negative items in multiple-choice test items. Doctoral dissertation, SUN 4 Buffalo, DAI-Univ. microfilms, n° 69-20, 512.
- THORNDIKE, R.L. (1971), Educational measurement for the Seventies, in Thorndike, R.L. (Ed.), *Educational Measurement*, American Council on Education, Washington.
- TRIVERS, R., 1955, *Educational Measurement*, New York, Mc Millan Co.
- VANDELDELDE, L. (1971), *Réflexions sur les QCM*, Bruxelles, ULB, document ronéotyp.
- WAHLSTROM, M. & BOERSMA, F. (1968), The influence of test-wisness upon achievement, *Educ. and Psychol. Measmt.*, vol. 28, 413-420.
- WEST, P. (1923), A critical study of the right minus wrong method, *Jour. of Educ. Research*, vol. 8, 1-9.
- WILBUR, P. (1970), Positional response set among high school students in multiple-choice tests, *Journ. of Educ. Measmt.*, vol. 7, 161-163.
- WILLIAMSON, M.L. & HOPKINS, K.D. (1967), The use of 'none of these' vs homogeneous alternatives on multiple-choice tests : Experimental reliability and validity comparisons, *Jour. Educ. Measmt.*, 4, 53-58.
- WILMUT, J. (1975b), *Selecting Objectives Test Items*, Associated Examining Board, Aldershot.
- WOOD, R. (1973) A technical report on the multiple choice tests, London GCE Board 1971 and 1972, Unpublished document, University of London, School Examinations Department, London.
- WOOD, R. (1974), Guessing on objective type test items, *School Science*, vol. 56, 179-180.
- WOOD, R. (1976), Inhibiting blind guessing : the effect of instructions, *Journ. of Educ. Measmt.*, vol. 13, 297-307.
- WOOD, R. (1977) Multiple-choice : A state of the art report, in CHOPPIN and POSTLETHWAITE (Eds), *Evaluation in Education International Progress*, Pergamon, Oxford.

DANS LA MEME COLLECTION :
EDUCATION 2000 dirigée par G. DE LANDSHEERE

**LA PEDAGOGIE PALEOLITHIQUE
 OU PREHISTOIRE DE LA CONTESTATION**
 de H. BENJAMIN. TRADUCTION de G. DE LANDSHEERE

PSYCHOLOGIE DE L'ENFANT ET DE L'ADOLESCENT
 de J.A. RONDAL et de F. HOTYAT - 1^{re} édition

**EVALUATION CONTINUE ET EXAMENS
 PRECIS DE DOCIMOLOGIE**
 de G. DE LANDSHEERE - 5^e édition

TECHNOLOGIE EDUCATIVE ET AUDIO-VISUEL
 de T. DECAIGNY - 3^e édition

LE TEST DE CLOSURE
 de G. DE LANDSHEERE - 4^e édition

COMMUNICATION AUDIO-VISUELLE ET PEDAGOGIE
 de T. DECAIGNY - 2^e édition

CONSTRUIRE UN COURS PROGRAMME
 de D. LECLERCQ, J. DONNAY, R. DE BAL et P. LAMBRECHT - 2^e édition

COMMENT MESURER LA LISIBILITE
 de G. HENRY - 2^e édition revue et corrigée

LA RELATIVITE EDUCATIONNELLE
 de A. CLAUSSE

L'ETUDE EN EQUIPES
 de G. POIRIER

CONCEPTS ET METHODES DE LA STATISTIQUE. Vol. 1 et 2
 de L. D'HAINAUT

**LA MISE EN PLACE DES POLITIQUES EDUCATIVES :
 ROLE ET METHODOLOGIE DE LA CARTE SCOLAIRE**
 de J. HALLAK

DES FINS AUX OBJECTIFS DE L'EDUCATION
 de L. D'HAINAUT - 3^e édition revue et augmentée

INTRODUCTION A LA PSYCHOPEDAGOGIE DE L'EXPRESSION. Vol. 1 et 2
 de J. FABRY

COMMENT SE DOCUMENTER
 de J.-E. HUMBLET

CARACTERISTIQUES INDIVIDUELLES ET APPRENTISSAGES SCOLAIRES
 de B. S. BLOOM

LE GROUPEMENT DES ELEVES EN EDUCATION
 de A. YATES

LES COMPORTEMENTS NON VERBAUX DE L'ENSEIGNANT
 de G. DE LANDSHEERE et A. DELCHAMBRE

LE GRAPHISME PHONETIQUE
 de V.A. GLADIC

**ANALYSE ET REGULATION DES SYSTEMES EDUCATIFS,
 UN CADRE CONCEPTUEL**
 de L. D'HAINAUT

**ELEMENTS DE PSYCHOLOGIE :
 UNE INTRODUCTION A LA PSYCHOLOGIE GENERALE**
 de J.A. RONDAL

EDUQUER LES PARENTS
 de J.-P. POURTOIS

DROGUES, UN PROGRAMME D'EDUCATION POUR LA SANTE
 de Ernst SERVAIS

POURQUOI LOGO DANS UN CONTEXTE EDUCATIF ?
 de J.-L. HARDY

LES JEUNES, L'ECONOMIE ET LA CONSOMMATION
 Sous la direction de J.-M. ALBERTINI

ENTRER A L'UNIVERSITE
 de J.-P. POURTOIS

**AUTONOMIE ET CONDITIONNEMENT CHEZ
 L'ENFANT ET L'ADOLESCENT**
 de R. GUBBELS

**UNE METHODE D'ECRITURE DES
 DIDACTICIELS - PILOT**
 de C. DEPOVER

