

Règles d'élaboration d'une évaluation par Questions à Choix Multiple

Joël LECHEVALLIER¹

Préambule

Les questions à choix multiple (QCM) sont une méthode à la fois fiable, objective, valide et rentable pour mesurer la connaissance. Pourtant, les enseignants qui ont l'expérience de ce type d'évaluation la critiquent volontiers pour plusieurs raisons : les questions sont difficiles à rédiger, le champ de connaissances exploré est restreint, les leurres sont complexes à concevoir, les étudiants répondent de mémoire à des questions mises à disposition dans une banque sans en comprendre la réponse... De leur côté, les candidats examinés stigmatisent (le plus souvent à juste titre) le caractère stupide des leurres, la complexité des questions « tirant dans les coins », l'ambiguïté de certaines propositions, ou leur caractère discutable.

Comme tout autre procédé d'évaluation la QCM doit être considérée comme un outil. De même que nous apprenons à nos internes que la pince de *de Bakey* est plus adaptée pour saisir un vaisseau qu'une pince à disséquer à griffes, nous devons admettre que les QCM sont un outil particulièrement adapté pour l'évaluation des connaissances fondamentales (savoir), alors que les tests à concordance de script seront plus adaptés pour l'évaluation du raisonnement (savoir comment) ou l'observation directe pour l'évaluation du comportement (savoir être).

Notre propos est donc de préciser les critères de rédaction et d'utilisation d'une QCM.

Construction d'une QCM

Principes directeurs

A aucun moment au moment de la rédaction d'une QCM, il ne faut perdre de vue notre objectif d'évaluation d'une connaissance factuelle. En pratique, cela signifie que l'on ne devra pas solliciter chez le candidat évalué d'autres compétences telles que résistance au stress, capacité de discernement, vigilance, raisonnement, intelligence stratégique... Non pas que ces qualités ne soient requises pour nos candidats mais parce que ce type d'évaluation n'est pas un outil adapté pour évaluer ces qualités.

Le champ de connaissances exploré doit être en rapport avec le niveau de compétence requis. En particulier, la concordance avec les objectifs d'apprentissages prédéfinis est une précaution élémentaire. Il est admis qu'un enseignant expert proche de la discipline doit être en mesure de répondre correctement aux questions.

Choix de la vignette

La vignette ou corps de la question est la première partie de la QCM. Elle situe le cadre et la thématique de la QCM.

La proposition ne doit pas comporter plusieurs affirmations qui, au lieu d'apporter de la précision génèrent de l'hésitation.

Cette vignette pourra être longue et détaillée (QCM riches), pour autant que sa richesse ne soit pas de nature à créer de l'ambiguïté ou du flou dans le sens de la question.

Idéalement, la formulation de la vignette doit être assez complète de sorte qu'un expert pourra être en mesure de donner la bonne réponse sans avoir à consulter les propositions.

¹ Joël LECHEVALLIER Clinique Chirurgicale Infantile. CHU de Rouen. 76031 ROUEN Cedex Joel.Lechevallier@chu-rouen.fr

Si la vignette est :

A propos de la péritonite appendiculaire quelles sont les propositions exactes ?

On notera que la thématique est très générale voire vague, de sorte que les propositions seront disparates. En revanche :

Quels sont les arguments cliniques en faveur d'une sténose hypertrophique du pylore ?

Cette formulation est beaucoup plus fermée et permettra de focaliser les propositions de réponse sur la même thématique

Choix des leurres

Nombre de leurres

Une bonne QCM doit en règle comporter 4 leurres et une seule bonne réponse, au moins pour une évaluation sanctionnante dont l'objectif est la validation d'une connaissance.

- Les QCM pour lesquels plusieurs bonnes réponses sont attendues doivent être utilisés avec parcimonie. En effet, il est inhabituel que, si trois bonnes réponses sont attendues, la valeur de chacune soit équivalente, l'une d'elle est souvent plus rare, moins réaliste, plus discutée... Dans une telle hypothèse, le candidat qui ne donne que deux bonnes réponses sur trois se trouve injustement pénalisé (0 points).

Si l'évaluation se fait dans le cadre d'un concours, l'objectif est différent. La discrimination va du meilleur candidat au plus mauvais. On souhaite départager chacun. L'usage du patron de réponse peut alors être remis en question.

Mais il faut savoir que la multiplication des bonnes réponses possibles dans un QCM diminue la valeur discriminante. Il est plus facile d'isoler un leurre, si on attend 4 bonnes réponses que 4 leurres si on attend une seule bonne réponse. On a vu proposer des QCM avec 0 ou 5 bonnes réponses ; il est évident que dans un tel cas la proposition d'un patron de réponse serait ridicule, mais en outre sans patron, très peu de candidats auront l'idée qu'un tel piège peut être tendu.

C'est ainsi que la première règle annoncée reste la plus sûre : **Une bonne réponse et 4 leurres**. Si l'on cède à la tentation de se libérer de cette règle, il ne faut pas fournir plus de 3 bonnes réponses (et donc 2 leurres) sur 5 propositions. Dans ce cas, il faudra **fournir un patron de réponse** dans la vignette : « quels sont les deux propositions exactes ? », « Quels sont les deux diagnostics les plus pertinents ? »...

Qualité

Le choix de 4 leurres est une contrainte reconnue par tout enseignant s'étant déjà prêté à l'exercice de la rédaction de QCM.

- Le leurre doit être **réaliste** : son exclusion ne doit pas être évidente ou relever d'un travail de déduction plus qu'un appel aux connaissances.

Par exemple, dans la QCM suivante :

Parmi les régimes alimentaires suivants, lequel convient pour un nourrisson de 1 mois ?

A 6 biberons de lait 2ème âge de 120 g chacun

B 6 biberons de lait 1er âge de 120 g chacun

C 6 biberons de lait 1er âge avec adjonction de légumes mixés dans un biberon

D 4 biberons de lait 1er âge de 210 g chacun

E 4 biberons de lait 1er âge de 110 g + 100 g de viande

Un étudiant normalement intelligent, parfois aidé d'une petite expérience parentale, n'aura pas de mal à exclure le lait 2^{ème} âge pour un nourrisson de 1 mois, ou la viande, fut-elle achée dans le biberon !

- le leurre ne doit comprendre **qu'un message** au risque d'entraîner le doute ou la confusion.

Par exemple :

Le vaccin contre la rougeole :

A) *Est administré en France par voie buccale*

B) ...

Dans ce cas si la question porte sur le mode d'administration (buccale), sur la réglementation française, ou sur les deux.

3. Le leurre doit être autant que possible recruté dans le même domaine que la bonne réponse (options cliniques, options para cliniques, options thérapeutiques...).
4. Si des unités de mesures sont utilisées, il ne faut pas chercher à piéger le candidat par des unités différentes.
5. Enfin, mais c'est une évidence, le leurre doit être faux de façon indiscutable et consensuelle.

Syntaxe

Structure de la phrase

Les phrases doivent être aussi courtes que possible, dans la vignette, mais aussi dans les propositions de réponses.

Il faut veiller à ce que la structuration de chaque proposition de réponse soit identique pour une QCM donnée. Toute variation dans la forme ou dans l'usage de négation (cf ci-dessous) constituera des pièges ou des indicateurs pour l'étudiant. Il identifiera une proposition donnée non pas pour son contenu mais pour sa forme.

Formulation des phrases

L'usage de formes affirmatives doit être la règle. Toute autre formulation est de nature à créer une ambiguïté, voire un piège, ce qui n'est pas l'objectif de l'examen.

Ce qu'il ne faut pas faire

Il faut éviter les formulations ambiguës telles que les formes négatives ou a fortiori les doubles négations :

➤ *Dans quel cas la fièvre est absente ?*

➤ *Dans quel cas il n'y pas d'apyrexie ?*

On préférera

➤ *Dans quel cas la température est normale ?*

Vocabulaire

Champ lexical

Les mots utilisés ne doivent pas être ambigus ni appartenir à un champ lexical autre que celui qui est requis pour nos candidats.

Les termes techniques médicaux sont donc acceptés voire bienvenus pour peu qu'ils soient utilisés couramment dans notre spécialité. En revanche, les mots issus du dictionnaire *Larousse* susceptibles d'être ignorés d'un candidat dont le français n'est pas la langue maternelle doivent être évités.

Dans la QCM suivante...

La sténose du pylore :

A) *peut s'opérer par une voie d'abord transrectale*

B)...

... il apparaît que l'enseignant a cherché à piéger le candidat sur l'ambiguïté du mot « transrectal » plus qu'à évaluer sa connaissance.

Adverbes et formes adverbiales

Il est connu que la médecine se prête mal aux affirmations définitives et c'est l'une des raisons pour lesquelles il est complexe de concevoir une bonne QCM. C'est la raison pour laquelle l'usage des adverbes, faute d'épurer le style rédactionnel, est un recours tentant pour se protéger contre une contestation par l'érudit qui saura qu'il a été publié un jour un cas de... C'est ainsi que l'on voit souvent écrire : le plus souvent, rarement, exceptionnellement, toujours... Cet usage porte en lui, dans certains cas, la réponse :

Le vaccin contre la rougeole :

A)...

B)...

C) Nécessite toujours un rappel 1 an après

Dans ce cas, l'usage de « toujours » apparaît comme un « parapluie » de l'examineur ; l'étudiant malin pourra écarter cette réponse en se disant que « toujours » n'existe pas dans le vocabulaire médical.

Certains adverbes portent en eux une indication de fréquence. Il est donc important, si l'on souhaite y recourir de rappeler en début d'examen la valeur sémantique de chaque adverbe :

- Classiquement : > 90% des cas
- Typiquement, Habituellement : >60% des cas
- Souvent, Parfois, Peut... > 30% des cas
- Est associé à ..., est le traitement de choix..., peut être vu..., peut être causé par... sont autant de formes adverbiales qui ne portent pas en elle d'indication de fréquence significative.

Durée d'exposition à la QCM

Quel que soit le format de présentation utilisé (vidéo QCM ou QCM écrit), il faut prévoir un temps de composition pour chaque. En moyenne on retient le principe d'une QCM par minute. Mais il est possible de prévoir que certaines QCM portant sur des connaissances « basiques » doivent être achevées en 30 secondes et que d'autres pour lesquelles la lecture de la vignette ou des propositions de réponse est longue requièrent 3 minutes de réflexion.

Modes de pondération

Mode binaire

C'est le procédé le plus couramment utilisé :

- Bonne réponse = 1 point,
- Toute autre réponse = 0 point

Ce procédé a l'avantage de la simplicité d'utilisation. Il manque toutefois de subtilité, en particulier quand il y a plus d'une réponse parmi les 5 propositions. En effet les candidats se trouvent à égalité s'ils n'ont pas donné toutes les bonnes réponses, et on ne distinguera pas celui qui a donné 2 bonnes réponses sur 3 de celui qui n'en a donné aucune, voire fait 2 mauvaises réponses.

Au prorata

Dans cette formule le candidat totalise d'autant plus de points qu'il a fourni de bonnes réponses. Cela a déjà été souligné plus haut :

- Deux bonnes réponses (sur 2) = 1 point
- Une bonne réponse = ½ point
- Aucune bonne réponse = 0 point
- On peut aussi déterminer que certaines mauvaises réponses seront sanctionnées, (points négatifs).

C'est la procédure utilisée pour l'examen du collège :

1. question comportant une bonne réponse :
 - a. bonne réponse : 1 point
 - b. toute autre réponse : - 0,25 point
 - c. la note minimale à la question est 0
2. question comportant 2 bonnes réponses
 - a. chaque bonne réponse : 0,5 point
 - b. chaque mauvaise réponse : - 0,33
 - c. la note minimale à la question est 0

3. question comportant 3 bonnes réponses
 - a. chaque bonne réponse : 0,33 points
 - b. chaque mauvaise réponse 0,5 point
 - c. la note minimale à la question est 0

Pondération variable

Il est possible de ne pas donner la même pondération à chaque bonne réponse en fonction de l'importance que le jury lui donne. Cette règle peut s'appliquer

- au niveau de la Question : une QCM difficile ne comptera que pour 1 point ; une QCM portant sur un thème essentiel comptera pour 3 points
- au sein même de la question : le choix d'un leurre jugé inacceptable pourra provoquer un 0 à la question ; le choix d'une des deux bonnes réponses attendues pourra rapporter 0,75 point, alors que le choix de l'autre, jugée moins importante ne rapportera que 0,25 point.

Réutilisation des épreuves

Il est classique, en France, de laisser aux candidats le loisir de conserver à l'issue de l'épreuve la listes des questions posées. Cette pratique a des effets délétères à plusieurs titres :

- Les candidats créent des banques de QCM qui permettent aux générations qui les suivent un apprentissage avant l'épreuve forcément bénéfique. Parfois, on observe malheureusement que les étudiants apprennent par cœur la bonne réponse à les QCM incomprises.
- Les banques étudiantes « épuisent » les enseignants qui ne sont pas en mesure de produire pour chaque nouvel examen un nombre assez élevé de nouvelles QCM de bonne qualité.
- La sortie régulière de la même question fait bien-sûr perdre la valeur du taux de bonnes réponses si elle est attendue par les étudiants.

En pratique, il faut savoir que dans les pays anglo-saxons, il est formellement interdit de quitter l'épreuve avec un quelconque document (original ou recopié), sous peine de très sévères sanctions.

Une alternative est l'usage de vidéo-QCM, puisqu'il n'y a pas de trace papier.

Quelle que soit la procédure utilisée, il faut savoir que les parades restent possibles (un étudiant retient par cœur la première question, un autre la deuxième...). C'est dire que l'on est contraint de faire usage chaque année de QCM comportant au moins 60% de questions nouvelles, en sachant que le renouvellement peut porter sur un leurre une bonne proposition, la formulation de la vignette...

Questions d'ancrage

A contrario, il est utile voire nécessaire de faire usage chaque année d'un quota (10% à 30%) de questions d'ancrage déjà validées, déjà utilisées et pour lesquelles la performance des étudiants est connue. Cette pratique permet de mesurer de façon comparative les différentes cohortes de candidats. Si une cohorte a réalisé une performance faible, mais qu'elle a répondu comme les cohortes précédentes aux questions d'ancrage, c'est que l'épreuve était de niveau de difficulté supérieure à celle des années précédentes. Si en revanche sa performance était aussi plus faible que les années précédentes sur les questions d'ancrage, c'est que le niveau de la promotion était inférieur à celui de ses aînés.