



^b
**UNIVERSITÄT
BERN**

Medizinische Fakultät

Institut für Medizinische Lehre IML

**Abteilung für Assessment und
Evaluation AAE**

Mode d'emploi pour la création de questions et d'examens à choix multiples pour l'enseignement de la médecine

René Krebs

contact:
rene.krebs@iml.unibe.ch

download:
www.iml.unibe.ch

Table de matières

Qu'allez-vous apprendre dans ce mode d'emploi?	1
1. A quelles exigences les questions à choix multiple doivent-ils répondre?	2
2. Que vise le thème d'une question?	4
3. A quoi se prête quel type de question?	5
3.1 Choix unique positif parmi cinq réponses possibles (Type A+)	6
3.2 Choix unique négatif parmi cinq réponses possibles (Type A-)	9
3.3 Attribution (Type B)	10
3.4 Attribution élargie (Type R)	11
3.5 Choix d'un nombre imposé de meilleures réponses (Type PickN)	12
3.6 Quadruple décision vrai/faux (Type Kprime/K')	13
3.7 Relation de causalité (Type E)	15
4. Comment formuler les questions?	16
4.1 A quoi faut-il veiller lors de la formulation du tronc?	17
4.2 A quoi faut-il veiller lors de la formulation des réponses?	17
Test UTC	19
5. Comment vérifie-t-on les questions?	22
6. Comment compose-t-on un examen?	23
Sources, littérature	24
Annexe 1: Bases pour la réalisation de questions standardisées	25
Annexe 2: Formulaire pour la rédaction des questions	26
Annexe 3: Informations générales sur la manière de procéder	27
Annexe 4: Instructions pour répondre aux différents types de questions	28
Glossaire des termes techniques spécifiques aux examens	29

Berne, novembre 2004

© Institut für Medizinische Lehre IML
Abteilung für Assessment und Evaluation AAE
Inselspital 37a, 3010 Bern

Traduction française: Dr méd. Rainer W. Bielinski et Dr méd. Monique Savopol

Qu'allez-vous apprendre dans ce mode d'emploi?

Ce mode d'emploi montre

- **comment déterminer les thèmes d'items importants***
- **à quoi se prêtent les différents types d'items**
- **à quoi il convient de faire attention lors de la formulation des items**
- **comment tester les items nouvellement formulés**
- **à quoi il faut veiller lors de la composition d'un examen**

En vous conformant aux instructions figurant dans ce mode d'emploi lors de la conception d'items à choix multiple, vous testerez davantage que de simples connaissances spécifiques isolées et vous éviterez de nombreuses fautes de formulation.

Mais, c'est en forgeant qu'on devient forgeron. On recommandera donc aux nouveaux auteurs de questions de prendre le temps d'assister à un atelier d'au moins un jour pour élaborer et vérifier avec leurs collègues les items. Les auteurs devront être régulièrement informés des résultats de la révision de leurs questions de manière à pouvoir tirer les leçons de leurs erreurs. On n'oubliera pas non plus qu'il faut du temps pour bien faire les choses. Même des auteurs expérimentés comptent en moyenne une heure de travail pour la conception d'un seul item à choix multiple.

Le chapitre 1 décrit à quoi il faut veiller pour que les items à choix multiple contribuent vraiment à la validité (validity) et à la fiabilité (reliability) d'un examen. Les chapitres 2 - 5 forment le noyau de ce document et sont consacrés à l'élaboration concrète des items: on commence par choisir le thème et le type d'item qui convient, tout en précisant à ce stade déjà les exigences spécifiques liées à cet item. Suivent des remarques d'ordre général, destinées à assurer une formulation correcte et à éviter de suggérer involontairement la solution (cues). Et pour terminer cette partie, nous donnerons quelques explications sur le processus de révision. Le chapitre 6 décrit à quoi il faut faire attention en rassemblant des items pour former l'examen.

* Vous trouverez en pages 29/30 un glossaire de tous les termes techniques spécifiques aux examens utilisés dans ce document

1. A quelles exigences les questions à choix multiple doivent-ils répondre?

L'examen QCM en tant qu'instrument de contrôle de qualité

Nous partons du principe que l'examen QCM planifié a pour objectif de tester un aspect qualitatif essentiel de la formation des étudiants examinés, c'est-à-dire des connaissances spécifiques primordiales et la capacité de les mettre en pratique en situation concrète.

condition préalable: définition et pondération du contenu

Les conditions pour que ce contrôle puisse se faire de façon représentative incluent la définition aussi précise que possible des connaissances spécifiques, idéalement sous la forme d'un catalogue des objectifs d'apprentissage, et l'élaboration d'un blueprint, c'est-à-dire d'une liste des matières examinées pondérée.

condition essentielle: de bons items

En définitive, la validité et la fiabilité dépendent de la qualité des items. Ces derniers sont très probablement adaptés à une utilisation dans le cadre d'un examen s'ils remplissent les critères suivants:

Voilà comment assurer la validité des items ...

Un item contribue à la validité d'un examen quand

- **le thème de la question correspond aux exigences futures**, que cela soit dans la formation ou dans la pratique professionnelle

Il s'agit d'éviter aussi bien les sophismes que les trivialités. Cf. critères de pertinence possibles en page 4

- **le niveau d'exigence est adéquat**

Des items qui n'impliquent que la réactivation de notions isolées apprises par cœur ne peuvent pas engendrer un examen valable, si l'objectif primaire de l'examen est d'évaluer la compréhension et la capacité de mettre en pratique les connaissances. Tant qu'on testera le contenu tel qu'on le trouve dans les livres et les photocopiés, on n'examinera que des **connaissances** au premier degré dans l'échelle taxonomique.

La compréhension peut être évaluée en demandant aux candidats

- d'analyser, de mettre en relation et d'intégrer des informations,
- de transformer des informations (p.ex. tableau, graphique → langage),
- de saisir les rapports entre différentes informations,
- de tirer des conclusions.

La capacité de mise en pratique et d'évaluation peut être examinée en demandant aux candidats

- d'appliquer le savoir acquis à une nouvelle situation (abstraire, transposer, généraliser),
- d'utiliser le savoir acquis pour résoudre certains problèmes,
- d'apprécier, d'évaluer, de pondérer les informations (observations, résultats) et d'en soupeser les conséquences.

- il se **focalise** sur un **contenu ou un problème clairement défini** et qu'il se situe dans un **domaine bien circonscrit**

Toutes les réponses proposées doivent tomber dans une seule et même catégorie. Devoir évaluer si une proposition de cause d'un problème donné a plus de chances d'être correcte qu'une définition ou une mesure à prendre n'a aucun rapport avec la réalité de la pratique professionnelle.

- il existe **manifestement une solution bien meilleure que les autres**

Les questions touchant à des domaines controversés ne conviennent pas à la méthode QCM, à moins qu'on ne demande explicitement la position d'une école donnée.

... et voici comment assurer la fiabilité de la mesure des items

Un item contribue à la **fiabilité** de l'examen s'il est capable de différencier les candidats quant à leurs aptitudes dans la matière testée, en d'autres termes, s'il possède une bonne sélectivité. Ce critère est respecté quand

- sa difficulté est **adaptée au groupe cible**

Les items qui posent surtout problème sont ceux qui sont trop difficiles et pour lesquels même les bons candidats doivent deviner la réponse.

- il est **formulé de façon simple et claire**

Il s'agit de tester des connaissances et non des capacités linguistiques, ni le facteur chance par interprétation.

- il **ne contient pas d'indices suggérant la solution par mégarde** (cues)

Il s'agit de tester des connaissances et non l'expérience du candidat en matière d'examens QCM ou les qualités de spécialiste dans la «résolution d'énigmes».

Ces trois points ne sont pas seulement importants pour assurer une bonne fiabilité, mais constituent également des conditions indispensables pour que l'examen puisse mesurer les connaissances de façon valable. Le chapitre 4 est consacré aux aspects touchant à la formulation et à la façon d'éviter les «cues».

A côté de ces exigences générales, on retiendra encore **quelques exigences spécifiques**, que nous présenterons dans la partie consacrée à la description des différents types de questions. On remarquera aussi qu'il existe des **différences de qualité de mesure selon le type de question**. Nous y reviendrons également dans la description des différents types.

2. Que vise le thème d'une question?

du domaine thématique au thème de l'item

Les chapitres du «blueprint» et les objectifs d'apprentissage suggèrent des thèmes possibles pour la création d'items. Dans la formulation d'une question, il est nécessaire de se limiter à un aspect particulier du thème choisi, p.ex. la cause (la plus vrai-semblable) d'un problème ou la façon correcte ou la plus efficace de procéder pour trouver l'origine d'un problème ou la démarche correcte, resp. l'attitude qui a le plus de chances de régler un problème. Si on ne parvient à formuler une question que de la manière suivante «A propos de XY, laquelle des affirmations suivantes est correcte?», le domaine concerné a toutes les chances d'être trop étendu et/ou trop hétérogène pour fournir la base d'un bon item de QCM.

critères de pertinence

Les items pertinents concernent le plus souvent des aspects particuliers d'un thème donné

- auquel on est spécialement **souvent** confronté dans la situation ou l'activité concernées
- qui pourrait avoir des **conséquences graves en cas d'erreur**
- où les **fausses idées** sont **répandues**
- qui est **essentiel pour la compréhension de futurs domaines d'enseignement**

Vérifiez l'importance du thème envisagé en vous posant cette question: «Dans quelle mesure est-il important pour le candidat qu'il soit à même de résoudre ce problème de manière autonome et/ou qu'il sache répondre à cette question?»

expériences personnelles comme source d'inspiration

Partir de problèmes vécus peut aboutir à des items non seulement importants, mais aussi proches de la pratique quotidienne. Il faut néanmoins éviter de choisir des cas particuliers «intéressants».

utilisation d'ouvrages de référence

Le recours à des ouvrages de référence est important pour vérifier et documenter la justesse et la pertinence d'une question. Les ouvrages de référence peuvent aussi être très utiles pour trouver de bons distracteurs. On ne conseillera cependant pas leur utilisation comme source d'inspiration à la recherche de thèmes de questions. Cela résulte en effet très souvent en des questions par trop «académiques», dans le mauvais sens du terme.

critère de relation avec la pratique

Vérifiez que les thèmes envisagés ont un intérêt pratique en vous posant la question suivante: «Cette question va-t-elle se poser en ces termes dans la suite de la formation et/ou de l'activité professionnelle des candidats?».

recherche structurée d'affirmations sur un thème donné

Un thème choisi peut conduire spontanément à une bonne idée d'item. Il vaut souvent mieux ne pas se focaliser de façon trop rigide sur un type d'item précis, mais plutôt commencer par collecter quelques affirmations. Pour ce faire, une grille peut être utile (cf. Annexe 1 p. 25).

3. A quoi se prête quel type de question?

limitation à un petit nombre de types

Depuis qu'on a commencé à développer des examens écrits objectifs, vers la fin du 19^e siècle, de nombreux types d'items ont été créés. Pour les examens de QCM du National Board of Medical Examiners aux Etats-Unis, on avait initialement proposé un choix de plus de dix types¹, chacun étant désigné par une lettre de l'alphabet qui est encore utilisée de nos jours (également dans le présent mode d'emploi). On a cependant reconnu peu à peu qu'il était préférable, tant sur le plan de l'élaboration des questions que de celle des réponses et des mesures de qualité, de se limiter à un plus petit nombre de types.

questions de type «meilleure réponse possible» ...

On distingue deux types de base en fonction de la consigne donnée: le type «meilleure réponse possible» et le type «vrai/faux». Dans le groupe des types **«meilleure réponse possible»**, le nombre de réponses à donner est indiqué; en général, il n'y en a qu'une seule. Les réponses fausses, c'est-à-dire les déviateurs ou distracteurs, ne doivent pas forcément se démarquer sur un mode noir/blanc de la réponse correcte. Il est possible d'inciter les candidats à effectuer une démarche intellectuelle plus complexe et une appréciation des zones grises souvent plus proche de la pratique. Les types «meilleure réponse possible» ci-dessous seront présentés dans ce mode d'emploi:

Type	Consigne
A+	choix simple positif parmi cinq réponses à choix
A –	choix simple négatif parmi cinq réponses à choix
B	attribution de la réponse correcte parmi cinq réponses possibles à plusieurs questions
R	comme le type B, mais avec jusqu'à 26 réponses à choix (A-Z)
PickN	comme le type A+ ou le type R, mais avec un choix portant sur plus d'une «meilleure réponse possible»

... avant les questions de type «vrai/faux»

Pour les questions de type **vrai/faux**, les candidats doivent décider de répondre par oui ou par non à chaque proposition. Pour être admissible, chaque affirmation doit être absolument sans équivoque. Les types vrai/faux comportent le danger de tester uniquement des notions apprises par coeur.

Les types vrai/faux ci-dessous seront présentés dans ce mode d'emploi:

Type	Consigne
Kprime	Quadruple décision vrai/faux
E	Appréciation de deux affirmations et de leur relation de causalité



Les items du type meilleure réponse possible doivent clairement prévaloir dans un examen QCM.

3.1 Choix unique positif parmi cinq réponses possibles (Type A+)

définition Une question ou un début d'affirmation est suivi de cinq réponses ou fins de phrases à choix, parmi lesquelles il s'agit de choisir la **seule juste** ou la **meilleure**.

le tronc peut être constitué d'une phrase ... Les troncs des items peuvent être constitués d'une question ou d'une affirmation en une phrase. Ce genre d'item permet de tester des connaissances pratiques, mais il examine en règle générale surtout des connaissances factuelles.

... ou d'une vignette plus une question. Des items QCM peuvent et doivent aussi tester autant que possible la capacité d'interpréter et d'intégrer des informations et des connaissances théoriques dans une situation concrète. Ceci nécessite des troncs de questions en deux parties. Une première partie plus longue décrit un problème rencontré dans le cadre de l'activité professionnelle. Cette description d'un contexte ou d'un problème est appelée une vignette. Celle-ci est suivie d'une question nettement démarquée et formulée très brièvement.

Les vignettes décrivant des patients doivent contenir un ou plusieurs des éléments suivants, dans l'ordre indiqué:

1. Âge, sexe, évent. profession, origine/ethnie (*Un aubergiste de 45 ans ...*)
2. Lieu du traitement (*... est amené aux urgences ...*)
3. Motif de la consultation (*... en raison de violentes céphalées ...*)
4. Durée (*... ininterrompues depuis 2 jours.*)
5. Anamnèse, évent. anamnèse familiale (éléments principaux)
6. Status, examen clinique
7. Event. résultats des examens complémentaires
8. Event. traitement initial, évolution, etc.

La partie informative peut également comporter des tableaux et des illustrations (photos, radiographies, graphiques) ou des citations provenant de publications.

réponse juste et déviateurs/distracteurs Les cinq réponses à choix doivent être le plus homogènes possible, tant du point de vue du contenu que de la forme, et se rapporter uniformément au tronc. Les quatre distracteurs ne doivent pas nécessairement être totalement faux. La bonne réponse doit cependant se démarquer positivement et de manière très claire.

exemples 1. *Une femme de 60 ans éprouve des difficultés à sortir de son fauteuil et à se tenir droite. Elle n'a en revanche aucune peine à marcher sur sol plat et parvient à fléchir ses hanches sans aucune difficulté.*

Lequel des muscles suivants est le plus probablement déficient?

- (A) *m. grand fessier*
- (B) *m. moyen fessier*
- (C) *m. obturateur externe*
- (D) *m. obturateur interne*
- (E) *m. grand psoas*

Anatomie (2^e année)

Clé: A

2. Un homme de 65 ans, obèse, souffre de troubles respiratoires progressifs. La spirométrie donne les résultats suivants:

- diminution des volumes pulmonaires
- rapport normal entre volume maximal expiré à la 1^{ère} seconde et capacité vitale (FEV₁/VK)
- réduction sévère de la compliance pulmonaire

La gazométrie révèle une hypoxémie artérielle et une hypocapnie au travail

Quelle est la cause la plus vraisemblable de ce tableau clinique?

- (A) un emphysème pulmonaire obstructif
- (B) une fibrose interstitielle diffuse des poumons
- (C) une augmentation du débit circulatoire pulmonaire en raison d'un shunt artérioveineux
- (D) une obésité
- (E) une bronchite chronique

Physiopathologie (3^e année): P 81, R 0.20 *

Clé: B

3. Une étude de contrôle de cas fournit les résultats suivants:

	Maladie +	Maladie -
Exposition +	23	27
Exposition -	77	173

Comment calculer dans ce cas l'odds ratio?

- (A) $\frac{77}{23 + 77} = 0.77$
- (B) $\frac{173}{27 + 173} = 0.87$
- (C) $\frac{23}{23 + 27} = 0.46$
- (D) $\frac{23 / 27}{77 / 173} = 1.91$
- (E) $\frac{23 / (23 + 27)}{77 / (77 + 173)} = 5.00$

p.ex. SPM (3^e année): P 87, R 0.25

Clé: D

4. Le menu d'un banquet était composé d'un rôti de poulet, de pommes de terre rôties, de petits pois, d'un éclair au chocolat et d'un café. La majorité des convives présentent des nausées deux heures après le repas, se mettent à vomir et à se plaindre de douleurs abdominales.

Lequel des micro-organismes suivants a-t-on le plus de chances de trouver dans les analyses des aliments?

- (A) *Escherichia coli*
- (B) *Proteus mirabilis*
- (C) *Salmonella typhimurium*
- (D) *Staphylococcus aureus*
- (E) *Streptococcus faecalis*

Microbiologie (3^e année)

Clé: D

* P = % de réponses justes (difficulté de l'item), R = sélectivité de l'item (cf. Glossaire p. 29)

5. Un homme de 53 ans souffre depuis 2 jours d'une dyspnée progressive et d'une toux avec expectorations purulentes. Il s'agit d'un fumeur régulier d'un paquet de cigarettes par jour depuis 30 ans. Sa température est de 37.2° C. Le murmure vésiculaire est affaibli avec présence de quelques ronchi et sibilances. La numération leucocytaire est de 9000/mm³ (9.0 G/l) avec une répartition normale. La coloration Gram des expectorations montre de nombreux neutrophiles et des diplocoques Gram négatif. Les clichés du thorax révèlent une hyperinflation. Quel est le diagnostic le plus vraisemblable?
- (A) un asthme
 - (B) une bronchite
 - (C) des bronchectasies
 - (D) une embolie pulmonaire
 - (E) une pneumonie à streptocoques

Médecine interne (examen final)

Clé: B

6. Un patient de 47 ans se présente aux urgences après une chute à vélo. Il est agité et présente un intense foetor éthylique. Il se plaint de douleurs thoraciques dans la région de la base gauche et dans l'hypochondre gauche. Le score de Glasgow est à 15. Il n'y a aucune plaie apparente. La fréquence respiratoire est de 38, les pulsations sont à 140, la tension artérielle à 90/60 mmHg. Quelle est l'explication la plus vraisemblable du tableau hémodynamique?
- (A) un choc cardiogène
 - (B) une hypovolémie
 - (C) une embolie pulmonaire
 - (D) une conséquence de l'intoxication éthylique
 - (E) un réflex vagal post-traumatique

Chirurgie (examen final): P 88, R 0.25

Clé: B

7. Un homme de 64 ans se trouve à l'hôpital avec un emphysème pulmonaire au stade terminal. Ses proches constatent qu'il les reconnaît, qu'il est lucide et bien orienté, mais qu'il ne s'intéresse plus autant à leurs visites et qu'il semble très passif. Quelle est l'explication la plus probable du comportement de ce patient?
- (A) une évolution vers un délire
 - (B) une évolution vers un trouble psychotique
 - (C) une péjoration de troubles de la personnalité
 - (D) un excès de sédation
 - (E) un repli sur soi

Médecine psychosociale (3^e année)

Clé: E

adéquation

La question de type A+ constitue encore et toujours le standard des questionnaires à choix multiple. Elle a largement fait ses preuves au niveau international dans de nombreux domaines, notamment sur le plan de la technique de mesure et devrait prévaloir quantitativement dans chaque examen QCM.

précisions importantes pour la formulation

Vous trouverez des instructions détaillées pour la formulation de ce type standard au chapitre 4 à partir de la page 16.

3.2 Choix unique négatif parmi cinq réponses possibles (Type A-)

définition

Une question ou un début d'affirmation est suivi de cinq réponses ou fins de phrases à choix, parmi lesquelles il s'agit de choisir **l'exception** ou celle qui **est la moins appropriée**.

exemples

1. *Lequel des antibiotiques suivants **ne doit pas** être administré à un enfant âgé de 18 mois souffrant d'une otite moyenne aiguë?*

- (A) *l'amoxicilline*
- (B) *le céfaclor*
- (C) *le co-trimoxazole*
- (D) *la doxycycline*
- (E) *l'érythromycine*

Pédiatrie (Examen final): P 81, R 0.25

Clé: D

2. *Laquelle des modifications suivantes **n'est pas la conséquence** d'une sténose mitrale?*

- (A) *une hypertrophie de l'oreillette gauche*
- (B) *une hypertrophie du ventricule gauche*
- (C) *une hypertrophie de la paroi du ventricule droit*
- (D) *une stase pulmonaire chronique*
- (E) *des dépôts d'hémosidérine dans les poumons*

Pathologie (3^e année) P 87, R 0.30

Clé: B

adéquation

Ce type d'item est certainement indiqué dans les rares cas où la connaissance d'une exception est essentielle, mais il est généralement utilisé d'une manière tout à fait différente pour tester les connaissances relatives aux quatre réponses positives. La réponse «juste» ne représente que le «déchet» des solutions proposées. S'il s'agissait d'une décision de type noir/blanc, il serait plus logique de l'examiner au moyen d'un item K' (ch. 3.6). Posez-vous cependant la question de savoir si la connaissance de l'événement XY le plus significatif, le plus probable, le plus dangereux n'est pas plus important et plus réaliste, ce qui justifierait plutôt le recours à un item de Type A+.

Sur le plan de la difficulté et de la sélectivité, les items A- s'avèrent en moyenne équivalents aux A+, mais du point de vue de la validité, un examen ne devrait pas comporter des items A- en grand nombre.

précisions importantes pour la formulation

- La négation sera mise en évidence par **caractères gras** ou soulignés.
- Toutes les réponses doivent être formulées positivement pour éviter de soumettre des doubles négations. C'est pourquoi, la réponse «Aucune des réponses proposées» ne peut pas être admise ici comme dernier choix.

3.3 Attribution (Type B)

définition

Dans le type B, cinq réponses à choix, désignées par les lettres A-E, sont proposées au début de la question. Deux à cinq questions ou affirmations au maximum suivent, souvent introduites par une phrase de liaison qui précise la mission d'attribution. A chaque question correspond **une seule réponse correcte** ou la **meilleure réponse possible**. Chaque réponse peut être correcte à plusieurs reprises.

Il s'agit donc en réalité d'une série de questions de type A+ associées à la même série de réponses possibles.

exemple

- (A) le captopril
- (B) le diazoxide
- (C) le dinitrate d'isosorbide
- (D) la nifédipine
- (E) le propranolol

Quelle substance inhibe

1. la libération de la rénine?
(A) (B) (C) (D) (E) Clé: E
2. la dégradation de la bradykinine?
(A) (B) (C) (D) (E) Clé: A
3. l'ouverture des canaux calciques?
(A) (B) (C) (D) (E) Clé: D

Pharmacologie (3^e année): 1. P 61, R 0.40; 2. P 90, R 0.51; 3. P 92, R 0.40

adéquation

Dans la forme représentée ici, le type B est aussi bien adapté que le type A+ du point de vue de la validité que de la fiabilité. C'est donc sur la base de l'étendue et de l'importance du thème sous-jacent qu'il faut décider si un item isolé de type A est suffisant ou si un complexe de type B avec plusieurs items fera mieux l'affaire. Il est par contre déconseillé de proposer des «questions» sous forme de définitions ou de descriptions de caractéristiques auxquelles il faut attribuer l'élément correspondant en guise de «réponse». Ce style de questions mots croisés ne teste que la connaissance de définitions et de faits au premier degré.

précisions importantes pour la formulation

- Une combinaison de questions, dont le contenu est inhomogène et pour lesquelles certaines réponses peuvent être écartées d'emblée, est à proscrire.
- Le fait de pouvoir attribuer une réponse à une question ne doit pas faire écarter d'emblée cette réponse pour les autres questions.
- En règle générale, on ne devrait pas associer plus de trois questions à un complexe B, si l'on veut éviter que le thème abordé par la question supplante les autres thèmes de l'examen. De plus, on amplifie le danger inhérent aux deux points précédents, ce qui augmente la probabilité de tomber juste en répondant au hasard et réduit ainsi la fiabilité de la mesure.

3.4 Attribution élargie (Type R)

définition

La structure est fondamentalement la même que celle du type B, si ce n'est que la liste des réponses peut comporter jusqu'à 26 réponses (A-Z).

exemple

(A) Adénovirus	(L) Haemophilus influenzae
(B) Aspergillus fumigatus	(M) Histoplasma capsulatum
(C) Bacillus anthracis	(N) Mycobacterium tuberculosis
(D) Candida albicans	(O) Mycoplasma pneumoniae
(E) Chlamydia psittaci	(P) Neisseria gonorrhoeae
(F) Coccidioides immitis	(Q) Neisseria meningitidis
(G) Coronavirus	(R) Pneumocystis carinii
(H) Corynebacterium diphtheriae	(S) Rhinovirus
(I) Coxiella burnetii	(T) Streptococcus pneumoniae
(J) Coxsackievirus	(U) Streptococcus pyogenes
(K) Epstein-Barr Virus	(groupe A)

Veillez indiquer pour chacun des patients suivants présentant un état fébrile le microorganisme qui a le plus de chances d'être l'agent responsable de la maladie décrite.

- 1. Un enfant de 7 ans présente un état hautement fébrile et des douleurs pharyngées. Le pharynx est hyperémié, l'amygdale droite tuméfiée est recouverte d'un exsudat mucoïde blanchâtre. Il y a une adénopathie sous-angulomaxillaire droite douloureuse. La culture du frottis sur plaque d'agar montre de nombreuses petites colonies β -hémolytiques inhibées par la bacitracine. Clé: U*
- 2. Un patient de 18 ans souffre depuis une semaine d'un état fébrile avec des douleurs pharyngées et un malaise général. Les amygdales sont tuméfiées et recouvertes d'un exsudat. On observe aussi la présence d'adénopathies cervicales diffuses et d'une splénomégalie. On trouve une lymphocytose avec des lymphocytes atypiques. Le test pour les anticorps hétérophiles est positif Clé: K*

Microbiologie: exemple traduit de Case, Swanson (1998), p. 77

adéquation

Comparé à d'autres types de questions, le type R a donné de bons résultats en termes de technique de mesure dans différentes études. On espère que, devant une liste de réponses suffisamment longue, les candidats s'apercevront qu'il est peu économique d'y rechercher la solution juste et qu'ils la déduiront d'eux-mêmes avant de la rechercher de façon ciblée dans la liste (un peu comme pour des questions à réponse libre).

précisions importantes pour la formulation

- On commence par créer une liste de réponses correspondant à un thème choisi. Des plaintes d'ordre général, non spécifiques, telles que douleurs abdominales, céphalées, lombalgies, état fébrile (cf. exemple ci-dessus), constituent d'excellents thèmes. Les listes de réponses pourraient alors p.ex. contenir des structures anatomiques, des diagnostics, des examens complémentaires ou des médicaments. On procède ensuite seulement à la formulation des questions.
- Les réponses doivent être homogènes quant à leur contenu et devraient idéalement être formulées sous forme de mots isolés, exceptionnellement de courtes phrases. On peut aussi imaginer des zones désignées sur une image ou un graphique.
- Les réponses doivent être disposées par ordre alphabétique ou logique.

3.5 Choix d'un nombre imposé de meilleures réponses (Type PickN)

définition

Ce type peut être construit comme le type A+ avec une liste de réponses plus longue ou comme le type R. La principale différence par rapport à ces deux types réside dans le fait que les candidats doivent choisir un nombre donné de réponses (2 - 5).

exemple

- | | |
|----------------------------|------------------------------------|
| (A) fer | (G) Vitamine B6 |
| (B) fluorure | (H) Vitamine B12 (cyanocobalamine) |
| (C) acide folique | (I) Vitamine C |
| (D) calcium | (J) Vitamine D |
| (E) Vitamine A | (K) Vitamine E |
| (F) Vitamine B1 (thiamine) | |

Choisissez pour chaque enfant les suppléments vitaminés et électrolytiques appropriés

1. Un nourrisson de 1 mois est amené en contrôle chez le médecin. Il est exclusivement nourri par allaitement et l'examen est parfaitement normal lors de la consultation. (Choisissez 2 compléments alimentaires)
Clé: B,J
2. Une fillette de 6 ans souffre de mucoviscidose. Elle ne prend aucun médicament. (Choisissez 3 compléments alimentaires)
Clé: E,J,K

Pédiatrie: exemple traduit de Case, Swanson (1998), p.100

adéquation

Ce type de questions convient à des problèmes pour lesquels il existe plusieurs options importantes, qui se distinguent nettement des autres, comme les réponses B et H dans la représentation qui suit.

F C I E D A G H B

Choix le plus mauvais

Choix le meilleur possible

précisions importantes pour la formulation

- Les règles de formulation sont les mêmes que celles qui s'appliquent dans le Type R.
- Le Type PickN ne devrait s'utiliser que si l'on se réfère à une évaluation en zone grise (tests ou examens les plus importants, diagnostics les plus probables). Pour de simples décisions vrai/faux, le type Kprime convient mieux (p. 13).

évaluation

Des études ont montré que lorsqu'on demande plus de deux réponses correctes par question, il peut s'avérer judicieux d'utiliser une échelle de points (comme dans le type Kprime). Cela permettra de réduire la difficulté statistique de ce type de questions et d'en améliorer la sélectivité.

p.ex. un demi-point pour: 2 réponses correctes sur 3
3 réponses correctes sur 4
3 réponses correctes sur 5

3.6 Quadruple vrai/faux (Type K', i.e. Kprime)

définition

A une question ou une affirmation incomplète suivent quatre réponses (ou fins de phrase), pour lesquelles il faut décider si elles sont **vraies** ou **fausses**. Pour obtenir un point, les quatre décisions doivent être justes (PE). Il s'est avéré très utile d'attribuer un demi-point (DP) à un candidat ayant donné trois bonnes réponses.^{2,3}

exemples

1. L'enrichissement de l'air inspiré par 5 % CO₂ entraîne chez un sujet en bonne santé:

- (A) une hypoxémie artérielle
- (B) une augmentation de la ventilation
- (C) une augmentation de la circulation cérébrale
- (D) un abaissement du bicarbonate plasmatique

Physiopathologie (3^e année): PE P 46, R 0.32
DP P 65, R 0.33

Clé: -++-

PE = 1 point pour 4 réponses correctes,

0 points pour moins de 4 bonnes réponses

DP = 1 point pour 4 réponses correctes, 1/2 point pour 3 bonnes réponses

2. Vous lisez la radiographie du thorax d'un patient chez qui vous suspectez un emphysème.

Qu'est-ce qui parle en faveur de ce diagnostic de suspicion?

- (A) une augmentation de l'espace rétrosternal
- (B) des sinus costo-diaphragmatiques émoussés
- (C) des côtes horizontales
- (D) une augmentation de la vascularisation pulmonaire en périphérie

Chirurgie (examen final): PE P 66, R 0.23
DP P 80, R 0.27

Clé: +++-

adéquation

Le type K' est indiqué, du point de vue du contenu, lorsqu'il s'agit de décisions qui dépendent de plusieurs aspects ou d'un problème dont la solution peut nécessiter plusieurs éléments d'information. Toutes les réponses doivent pouvoir être évaluées sur le mode noir/blanc. On n'abusera pas du type K' en rassemblant dans un item des affirmations totalement hétérogènes, portant sur un thème très étendu.

Du point de vue de la technique de mesure, les items K' sont un peu plus souvent sources de problèmes que les types A et B et doivent être éliminées en raison du non-fonctionnement d'une seule réponse partielle. En termes de sélectivité pourtant, les items K' sont équivalents aux types A et B, lorsqu'on n'utilise que la méthode PE attribuant 1 point aux 4 réponses correctes. Et avec la méthode DP, qui attribue 0.5 point lorsque 3 réponses partielles sur 4 sont correctes, la sélectivité tend même à être meilleure. Elles sont nettement plus difficiles dans la forme «tout ou rien» que les questions à choix unique et ne sont donc guère appréciées des candidats. La version DP améliore ces deux aspects.^{2,3}

Le type K n'existe plus!

Dans le type K cinq solutions étaient proposées en réponse à quatre affirmations 1-4

- (A) 1+2+3
- (B) 1+3
- (C) 2+4
- (D) seulement 4
- (E) les quatre

La sélectivité de ce type est bien inférieure à celle du type Kprime et ne devrait plus être utilisé. Si une affirmation est identifiée comme fautive, trois solutions proposées sont éliminées, ce qui suffit déjà à augmenter le facteur chance à 50 %.³

Précisions importantes pour la formulation

- La formulation du **tronc doit rester ouverte quant au nombre d'affirmations correctes sur les quatre**.

Autrefois ce critère était rempli par des formulations peu agréables à lire, du genre:

'Quelle/s est/sont la/les relation/s qui s'applique/nt?'

Actuellement nous mentionnons tout d'abord dans les instructions aux candidats que 0-4 affirmations peuvent être justes, indépendamment que la formulation soit au singulier ou au pluriel. Nous utilisons d'autre part des affirmations quantitativement aussi neutres que possible, telles que:

'Pour cet XY, il est juste que: ... on parle de: ... il est approprié de:'

ou une question quantitativement neutre, telle que:

'Qu'est-ce qui fait partie des XY?', *'Qu'est-ce qui parle en faveur d'un XY?'*

'Qu'est-ce qui entre en ligne de compte comme cause du trouble présent?',

'Qu'est-ce qui est approprié comme mesure?', *'Quand un XY est-il indiqué?'*

Si un tronc approprié ne peut être formulé que de la manière suivante:

'Évaluez les affirmations suivantes à propos de XY:'

on vérifiera de façon critique si le contenu de la question n'est pas trop hétérogène.

- Le **tronc doit toujours** être formulé **positivement**; il est aussi conseillé d'éviter les négations dans les réponses.

Les doubles négations (il est faux de dire qu'il est faux...) peuvent induire en erreur et perturber l'objectif de la mesure des connaissances.

- **Chaque affirmation** doit être **clairement juste ou fausse**.

- **Ne jamais fusionner deux affirmations en une seule** (p.ex. une affirmation avec sa motivation).

Il peut devenir difficile de comprendre ce qu'il faut réellement évaluer.

- **Chaque réponse** doit être **indépendante de toutes les autres**. On évitera notamment des réponses s'excluant mutuellement.

- **Des termes vagues**, tels que «habituellement», «fréquemment», «souvent», «associé à» sont à **éviter** dans la mesure du possible.

Les études ont montré que de tels termes sont interprétés de manière très variable en termes de quantification. Le fait d'approuver une affirmation aussi vague ne dépend donc plus uniquement des connaissances. 4,5

- **Les affirmations justes et fausses** devraient être **en nombre plus ou moins égal**.

Les auteurs ont clairement tendance à vouloir transmettre encore des connaissances au cours de l'examen et donc à préférer les affirmations correctes. Comme la majorité des candidats semble s'y attendre et juge préférentiellement les affirmations comme étant justes en cas d'ignorance, elles ont une sélectivité plutôt moins bonne que les affirmations fausses.

3.7 Relation de causalité (Type E)

définition

Deux affirmations sont reliées par le terme «parce que». Il s'agit d'évaluer si les affirmations sont **vraies** ou **fausses** sans se pré-occuper de la relation entre elles. Si les deux affirmations sont vraies, il faut ensuite vérifier s'il existe ou non une relation de causalité entre elles. Cela donne cinq réponses possibles:

- (A) + parce que + Les 2 affirmations sont vraies, la relation de causalité est vraie.
- (B) +/+ Les 2 affirmations sont vraies, la relation de causalité est fausse.
- (C) +/- La 1^{ère} affirmation est vraie, la 2^e est fausse.
- (D) -/+ La 1^{ère} affirmation est fausse, la 2^e est vraie.
- (E) -/- Les deux affirmations sont fausses.

exemple

La diminution des taux de 2,3-DPG dans les érythrocytes favorise l'oxygénation tissulaire

parce que

la courbe de dissociation de l'hémoglobine est déplacée vers la droite en cas d'abaissement du 2,3-DPG érythrocytaire.

(A)	(B)	(C)	(D)	(E)
+parce que+	+/+	+/-	-/+	-/-

Physiopathologie (3^e année): P 86, R 0.32

Clé: E

adéquation

Du point de vue du contenu, les questions de type E seraient indiquées dans des domaines où les relations de cause à effet jouent un grand rôle. Dans la pratique, les auteurs les utilisent souvent comme solution de secours lorsqu'ils ne trouvent pas quatre distracteurs plausibles pour un item de type A dans un thème donné.

Du point de vue de la technique de mesure, les items de Type E posent pas mal de problèmes. Les liens de causalité sont rarement complètement tranchés (p.ex. faut-il donner la réponse «vrai» à une relation de causalité, si la seconde affirmation n'est que l'**une** des causes parmi d'autres?). La décision entre les réponses A et B est plus souvent affaire d'appréciation que de connaissances spécifiques. Il est par ailleurs difficile de construire des items de type E, dans lesquels tout semble plausible lorsque l'une ou même les deux affirmations sont fausses. Les items E sont susceptibles de contenir des «cues» (indices), ce qui en limite la sélectivité.³



Les items de type E ne devraient donc être utilisés qu'avec parcimonie, si tant est qu'on les utilise.

précisions importantes pour la formulation

- Les deux affirmations doivent être refermées sur elles-mêmes et contenir toutes les informations nécessaires pour pouvoir être évaluées de manière totalement indépendante l'une de l'autre.
- Dans les items comportant deux affirmations vraies, la relation de causalité doit être clairement vraie ou fausse et ne pas dépendre de l'interprétation ou de l'appréciation des candidats.
- Qu'elle soit vraie ou fausse, la relation de causalité doit toujours sembler plausible pour le profane.
- Evitez la prédominance d'items E où tout est vrai, y compris la relation de causalité, ou d'items où seule la 1^{ère} affirmation est vraie.

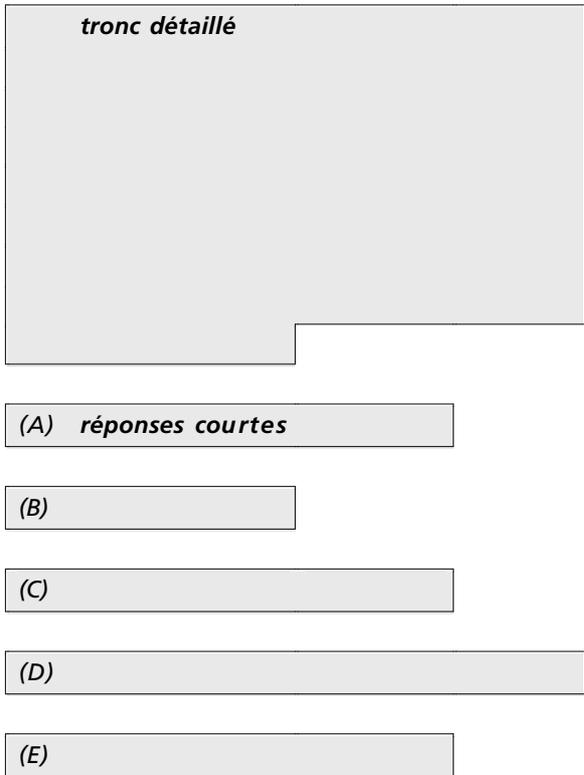
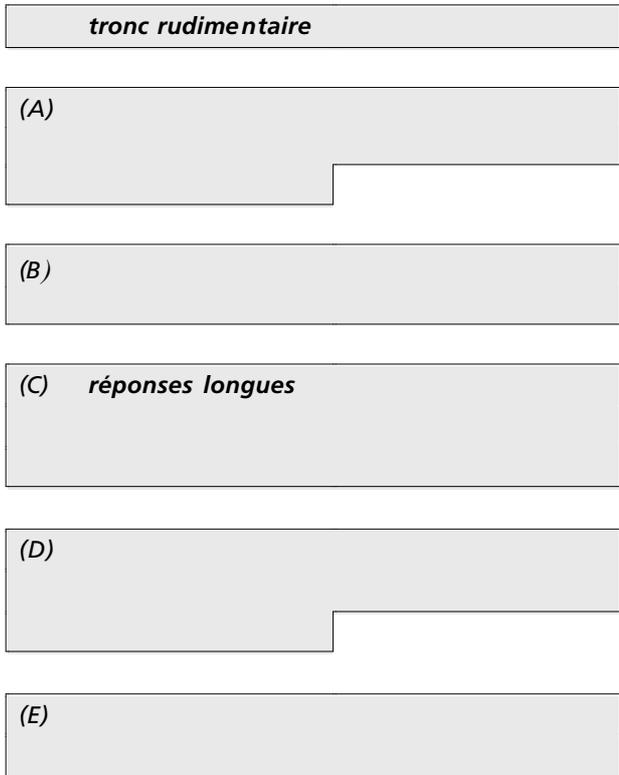
4. Comment formuler les questions?

Trois règles de base:

Si les items de QCM sont sensés tester davantage que des connaissances factuelles, il faut proposer des situations plus complexes, comportant plusieurs informations à interpréter et à intégrer. Les bons items nécessitent par conséquent souvent un **tronc plus détaillé**.

1. *tronc long, réponses courtes*

Dans les types «meilleure réponse possible», les **réponses** doivent souvent être appréciées l'une par rapport à l'autre. Ceci n'est pratiquement possible que si elles sont **courtes et précises**.

bonne structure	mauvaise structure
<p>tronc détaillé</p> 	<p>tronc rudimentaire</p> 

2. *formuler de façon simple et précise*

Les items doivent être formulés de façon **simple, claire** et dans un langage unanimement reconnu.

Ceci implique également de n'utiliser que des abréviations très courantes si l'on ne mentionne pas la version complète entre parenthèses.

3. *ne pas créer de questions pièges*

Les items **ne doivent ni être rendus artificiellement plus compliqués ni être conçus** volontairement **comme des questions pièges**.

La difficulté d'un item doit découler de la complexité du problème sous-jacent, des exigences cognitives (compréhension, capacité de synthèse, résolution du problème) et de la finesse de différenciation nécessaire (proximité des réponses à choix).

Il n'est pas correct de rendre artificiellement plus difficile des items sensés tester des connaissances de base ordinaires et cela entre certainement en contradiction avec les objectifs d'examen.

4.1 A quoi faut-il veiller lors de la formulation du tronc?

règles de formulation pour le tronc de l'item

Lors de la formulation du tronc, on veillera à ce que:

- l'affirmation/la question **contienne toutes les informations nécessaires** pour donner la réponse, de façon à ne pas avoir à en insérer d'autres dans les réponses à choix.
- d'une manière générale, elle ne contienne **pas d'informations superflues** ...
à moins qu'il s'agisse explicitement de tester la capacité de filtrer les informations importantes. A défaut, la lecture du texte superflu risque de prendre inutilement un temps précieux durant l'examen. Evitez tout particulièrement de tenter de transmettre encore des éléments d'enseignement pendant un examen servant à une sélection des candidats.
- on **puisse y répondre sans voir les réponses**
Si ce n'est pas le cas, demandez-vous si le contenu de l'item n'est pas trop hétérogène.
- dans les items à choix unique, **la formulation soit de préférence positive et qu'elle le soit toujours** dans les items Kprime
Une formulation négative dans le tronc des items K' entraîne presque inévitablement des doubles négations très perturbantes. Les items A- sont peu souhaitables du point de vue de la validité.

4.2 A quoi faut-il veiller lors de la formulation des réponses?

Pour qu'un item contribue à la validité et à la fiabilité de l'examen, la recherche des réponses justes devrait dépendre aussi exclusivement que possible des connaissances des candidats dans le domaine sensé être testé par l'item. Pour obtenir cela, les réponses à choix doivent répondre à un certain nombre de critères en termes de contenu, de formulation et de langage.

critères du contenu

- Toutes les réponses doivent toucher le même thème, donc **respecter l'homogénéité du contenu**.
Cette règle n'est de loin pas toujours respectée dans les items à thème mal focalisé et trop diffus.
- **Pour les distracteurs**, il faut des **notions claires**. Il peut par exemple s'agir de fausses idées largement répandues, de concepts erronés ou de vues dépassées. Il devrait néanmoins exister une relation clairement démontrée avec le thème de l'item.
Des distracteurs peu plausibles, simplistes ou même totalement absurdes peuvent être immédiatement éliminés par des candidats même faibles, ce qui augmente leurs chances de deviner la réponse correcte et diminue la sélectivité de l'item.
Pour réduire systématiquement la probabilité de deviner la bonne réponse, il est utile de proposer si possible cinq réponses à choix. Si l'on ne trouve véritablement que trois distracteurs plausibles, il est préférable de ne proposer que quatre réponses, plutôt que d'en inventer une absurde.

- Les distracteurs ne doivent pas forcément tous être totalement faux, mais la **réponse correcte** doit être **indubitablement la meilleure**. Plus la réponse correcte et les distracteurs sont proches en termes de contenu et de justesse, plus l'item va devenir difficile.
- Chaque réponse à choix doit être **aussi courte que possible** et ne doit contenir **qu'une seule affirmation**.
- On utilisera des **réponses qui se recourent** seulement si le problème sous-jacent l'exige et certainement pas pour augmenter artificiellement la difficulté d'un item.
- Des réponses du type **'toutes les réponses qui précèdent'** ou 'Les propositions B et C sont toutes deux correctes' **ne doivent pas être utilisées**.

S'il s'agit là de la bonne réponse prévue, cela signifie qu'il y a plusieurs bonnes réponses. S'il s'agit au contraire d'une réponse fautive, elle est «donnée» dès que l'un de ses éléments peut être identifié comme étant faux.

- **'aucune des réponses qui précèdent'** ne devrait être utilisé qu'**exceptionnellement** comme cinquième réponse à choix. On peut y recourir si elle teste des connaissances importantes ou lorsqu'on ne peut pas proposer la bonne réponse de manière explicite en raison d'une association trop forte (évent. même verbale) avec le tronc. Il est évident que la réponse 'aucune des réponses qui précèdent' devrait aussi être parfois une réponse fautive, mais cela pose souvent problème.

Lorsqu'on prévoit de proposer une réponse fautive de type 'aucune des réponses qui précèdent', la réponse correcte ne peut plus être simplement la meilleure possible dans la série. S'il existe une réponse encore meilleure, éventuellement plus différenciée que celle prévue par l'auteur, les très bons candidats vont choisir l'option 'aucune des réponses qui précèdent', ce qui fait perdre à l'item sa sélectivité.

critères formels

Pour les critères formels, il s'agit essentiellement d'**éviter de fournir des indices (cues) formels et verbaux**, qui permettraient aux candidats qui n'ont pas les connaissances requises, mais qui sont rompus à l'exercice des QCM, d'identifier la bonne réponse ou d'éliminer certaines mauvaises réponses, augmentant ainsi les chances de deviner la réponse correcte. Il faut par ailleurs **minimiser l'influence de certaines tendances systématiques dans les réponses** des candidats.

Testez votre capacité à résoudre les problèmes à l'aide du test UTC.

Pour vous aider à identifier les cues les plus importants et les plus fréquents, vous trouverez ci-dessous le test UTC que nous vous recommandons d'effectuer avant de continuer votre lecture.

Test UTC

Instruction: Veuillez cocher les réponses qui vous paraissent correctes dans les questions suivantes dans l'ordre proposé.

Conseil pratique: Ne vous laissez pas déconcerter par le contenu. Des questions à choix multiple peuvent parfois se résoudre sans connaissance particulière du sujet.

1. L'antère est une
 - (A) agglutination
 - (B) conglomérat
 - (C) produit de combustion
 - (D) produit de fission
 - (E) élément chimique

2. Lors de la fermentation de l'antère à l'aide du gore acide,
 - (A) il se produit une perte de chaleur
 - (B) on observe de l'antérite pour un pH se situant entre 2.8 et 3.2 après un léger réchauffement
 - (C) on constate une émanation d'Ogl₄
 - (D) de la gourmantine est produite
 - (E) il se forme du gorantocol

3. Vous voulez lubrier la gorantoxine à l'aide de votre ectateur. Pour ce faire, il faut que
 - (A) le module LTC de l'ectateur soit déclenché
 - (B) le module LTC de l'ectateur soit enclenché
 - (C) l'ectateur soit équipé d'un détecteur OC 2000+
 - (D) la fréquence Ω -reste constante pendant le lubriment
 - (E) la gorantoxine soit préalablement sternée

4. L'abréviation USL signifie en toutes lettres
 - (A) United States Laboratories
 - (B) Uniform Source Language
 - (C) Uniform Source Locator
 - (D) Uniform Starting Label
 - (E) Unique Spaceship Locator

5. Quelle est la caractéristique principale du SRC (syndrome de rigidité cognitive)?
 - (A) un taux de fermine augmenté dans le plasma
 - (B) des cauchemars post-prandiaux cycliques
 - (C) de fréquentes raideurs des muscles de la nuque
 - (D) une mobilité réduite dans le domaine cognitif
 - (E) une logo- et scriptorrhée chroniques

6. Pourquoi la probabilité de l'apparition d'une axosie est-elle augmentée chez les quéniches trigotes?
 - (A) Les quéniches trigotes ne sont jamais xéropes.
 - (B) La trigotie entraîne toujours une fébrilité accrue.
 - (C) Toutes les quéniches trigotes sont ésophymes.
 - (D) L'axosie provient exclusivement de la séquidité.
 - (E) Les quéniches trigotes sont le plus souvent sunimes.

7. L'antère ne peut pas être fermentée à l'aide du gore acide parce que un léger réchauffement est indispensable à la production de l'antérite.
 - (A) + parce que +
 - (B) +/+
 - (C) +/-
 - (D) -/+
 - (E) -/-

la solution de
l'énigme

sept cues fréquents
que vous devriez
éviter

Les distracteurs sont
au moins aussi impor-
tants que la réponse
correcte.

UTC est l'abréviation de 'Use The Cues'. Dans chacune des sept questions qui précèdent, on trouve un **indice formel ou verbal** typique. Il va sans dire qu'ils ne sautent pas toujours aux yeux, si non les auteurs les auraient repérés et donc certainement évités par eux-mêmes.

Vérifiez vos réponses et apprenez à reconnaître sept cues courants:

1. (A) Seul 'agglutination' correspond grammaticalement à la formulation du tronc.
 - **Toutes les réponses doivent correspondre au tronc du point de vue grammatical.**
 2. (B) C'est de loin la réponse la plus longue et la plus différenciée.
 - **Les distracteurs** devraient avoir **une longueur et une différenciation aussi proches que possible de celles de la réponse correcte.**
 3. (B) ou (A): deux fois 'le module LTC', soit 'déclenché', soit 'enclenché'
 - Éviter les **indications focalisant l'attention sur 2-3 réponses**. Le fait de répéter dans une seconde réponse des éléments clés cités dans la bonne réponse attire l'attention. Si une réponse fautive énonce pratiquement le contraire de la réponse correcte, on en déduira que l'une des deux devrait logiquement être la bonne.
 4. (C) Les réponses contiennent 3 fois le terme 'Uniform', 2 fois 'Source' et 2 fois 'Locator'. (C) contient les trois éléments.
 - **Les cues de convergence** sont à éviter. La réponse comportant le plus grand nombre d'éléments communs avec d'autres réponses a les plus grandes chances d'être la réponse correcte (stratégie dite de convergence).
- Les cues intégrés aux questions 1-4 proviennent du fait que la réponse correcte occupe fortement l'esprit des auteurs, si bien qu'ils accordent bien plus de soin à la formulation de cette dernière qu'à celle des distracteurs. De plus, ils tendent dans la recherche de mauvaises réponses à tourner continuellement autour de la réponse correcte. Evitez cette tendance!
5. (D) Le terme 'cognitif' se trouve dans le tronc et dans la réponse D.
 - **Les associations verbales entre tronc et bonne réponse** sont à éviter.
 6. (E) C'est la seule affirmation non absolue.
 - On évitera d'utiliser des **termes absolus**, tels que 'jamais', 'toujours', pour rendre des affirmations clairement fausses.
 7. (D) Des informations sur la bonne réponse sont évoquées dans la question 2.
 - **Les indices mutuels de solutions** sont aussi à éviter. On veillera surtout à ce type de cue au moment de composer l'examen. Ce danger est toutefois bien moindre si l'on évite de donner des informations utiles dans les items.

Deux autres erreurs formelles à éviter concernent l'ordre des réponses:

Éviter les cues dus à l'ordre des réponses et ...

- **L'ordre des réponses** devrait si possible **respecter une certaine logique**.

S'il existe dans le contenu des réponses une certaine logique hiérarchique, il est souhaitable de les disposer dans cet ordre (p.ex. dans la question concernant la cause la plus probable au problème, on fera suivre dans un ordre logique les termes tels que petit, bénin à grand, sévère, ou dans un ordre ascendant ou descendant s'il s'agit d'expressions numériques). S'il n'existe aucune logique interne et qu'il s'agit de réponses en un mot, l'ordre alphabétique est recommandé.

Les candidats ne connaissant pas la réponse correcte rechercheront aussi des indices dans la disposition des réponses pour les aider à trouver la solution.

... minimiser l'influence des tendances systématiques dans les réponses des candidats

- Il faut **éviter de positionner la réponse correcte préférentiellement en C ou D**.

Autrefois, les auteurs plaçaient volontiers la réponse correcte dans le milieu de la liste avec l'idée de mieux la «cacher». Il semble que cela se soit redit parmi les candidats, dont beaucoup privilégient les réponses du milieu lorsqu'ils doivent répondre au hasard. C'est la raison pour laquelle les auteurs placent aujourd'hui plus souvent la bonne réponse en D.

Pour éviter l'influence du choix en fonction de la position des réponses chez certains candidats, il convient de répartir le mieux possible les réponses correctes des items entre les positions A à E.

5. Comment vérifie-t-on les questions?

Le contenu et la forme des items de QCM doivent être vérifiés par l'auteur, mais aussi par des experts indépendants. Nous recommandons l'utilisation d'un formulaire standardisé (cf. Annexe 2). Cela garantit que toutes les informations nécessaires soient données et simplifie le travail de révision des nouveaux items que font les experts.

En tant qu'auteur, vous devez encore, après avoir rédigé un item:

Les auteurs contrôlent ...

- relire le texte, notamment en vous imaginant comment vous y répondriez vous-même
- **vérifier** sur la base des «points» du chapitre 1 de ce mode d'emploi **si toutes les règles de base ont été respectées**
- vérifier plus particulièrement l'absence de **cues**

... et documentent leurs items

- mentionner à quel **objectif ou domaine** (chapitre consacré au blueprint) l'item correspond et indiquer la **réponse correcte**
- préciser l'objet de l'item sous forme de **mots clés** et mentionner si possible une ou plusieurs **références de la littérature**, permettant de justifier la «réponse correcte»
Il vaut aussi la peine, le cas échéant, d'expliquer pourquoi les distracteurs ont été choisis et pourquoi ils sont faux.
- laisser ensuite reposer l'item une semaine, puis le reprendre «à tête reposée».

La révision de nouveaux items par des experts indépendants devrait se faire en plusieurs étapes.

Les experts en vérifient le contenu ...

- 1-2 experts de la branche concernée (contenu) en certifieront la **justesse** et l'**appartenance du thème au programme de formation**, indépendamment de la forme (cf. Formulaire de l'item dans l'annexe 2).

... et la forme

- Un expert en examens QCM devrait ensuite certifier les items exclusivement du point de vue de la **forme et de la langue**.

Les items sont remis avec les remarques des experts à l'organe de révision composé de représentants du domaine concerné.

La commission de l'examen décide

Les décisions d'**acceptation**, de **refus** ou de **modification** se prennent dans le cadre d'une séance de révision. L'appréciation de la **pertinence**, qui est très utile lors de la composition de l'examen, peut être calculée comme moyenne des appréciations individuelles ou être le fruit d'une discussion de consensus.

6. Comment compose-t-on un examen?

Pondération constante selon le contenu et les types

- Pour que les mesures soient comparables d'une session à l'autre, les examens devraient être non seulement équivalents en termes de contenu testé (blueprint), mais aussi sur le plan de la répartition des types d'items. Nous recommandons par conséquent de définir aussi une clé de répartition approximative pour les types d'items.

Instructions standardisées pour les réponses

- Le cahier de l'examen doit contenir une instruction standardisée décrivant la manière la plus sûre et la plus économique de répondre aux questions. Elle doit aussi informer les candidats sur la procédure d'évaluation des réponses et en particulier leur dire si les mauvaises réponses seront sanctionnées ou s'il vaut mieux répondre au hasard en cas d'incertitude (cf. Annexe 3).

Regroupement selon le type avec à chaque fois les instructions correspondantes

- Les items seront regroupés dans le cahier d'examen en fonction de leur type. Chaque groupe de types est introduit par une brève instruction sur la manière de procéder (cf. Annexe 4).

Il serait difficile et stressant pour les candidats à un examen de devoir constamment passer d'un type de question à un autre. De plus, il s'agirait d'une source d'erreurs indépendante des connaissances testées.

Entrée en matière avec un ou quelques items qui rompent la glace

- L'examen devrait commencer par 1-2 items faciles (pour dissiper l'angoisse des candidats).
Chaque examen constitue une situation de stress. Pour pouvoir véritablement tester des connaissances (et non la résistance au stress), il est utile de commencer l'examen avec quelques questions «gentilles».

Répartition équilibrée des réponses

- La répartition des réponses correctes mentionnées de A à E devrait être équilibrée sur l'ensemble de l'examen. Il est également préférable d'éviter des suites de cinq questions ou plus ayant la même réponse correcte.

Cela permet d'une part d'éviter l'influence éventuelle de certaines tendances chez les candidats. (Une tendance bien connue est celle du milieu de la liste). Et cela permet d'autre part d'éviter que les candidats soient déséquilibrés par une longue suite de réponses identiques suscitant la confusion dans leur esprit et les incitant à changer de réponse à leur détriment.

Sources

Les indications données dans le présent mode d'emploi sont basées sur une longue expérience personnelle dans le développement, la révision et l'évaluation des examens QCM dans le domaine de l'enseignement et de la formation des médecins en Suisse, ainsi que dans d'autres domaines d'application de cette forme d'examen. Elles s'appuient également sur une série de publications internationales traitant du thème «Développement d'items à choix multiple » (6,7,8,9,10). Nous aimerions attirer en particulier l'attention sur l'excellent manuel de Case et Swanson⁶ que l'on peut télécharger sur Internet: www.nbme.org/about/itemwriting.asp

Littérature

- 1 Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine. Philadelphia: Lea & Febinger, 1961
- 2 Krebs R. The Swiss way to score multiple true-false items: theoretical and empirical evidence. In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg, eds. Advances in Medical Education. Dordrecht: Kluwer Academic Publishers, 1997:158-61
- 3 Itten S, Krebs R. Messqualität der verschiedenen MC-Itemtypen in den beiden Vorprüfungen des Medizinstudiums an der Universität Bern. AAE/IML, Forschungsbericht 1997/2, 24 S.
- 4 Case SM. The use of imprecise terms in examination questions: how frequent is frequently? Acad Med 1994;69(suppl):4-6
- 5 Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Med Educ 1998;32:343-50
- 6 Case S, Swanson DB. Constructing Written Test Questions For the Basic and Clinical Sciences. 2nd ed. Philadelphia: National Board of Medical Examiners, 1998
- 7 Gronlund NE. How to construct achievement tests. 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 1987
- 8 Linn RL, Gronlund NE. Measurement and assessment in teaching. 8th rev. ed. Englewood Cliffs, NJ: Prentice-Hall, 1999
- 9 Ellsworth AR, Dunnell P, Duell OK. Multiple-choice test items: What are textbook authors telling teachers? J Educ Res 1990;83:289-93
- 10 Aiken LR. Testing with multiple-choice items. J Res Dev Educ 1987; 20:44-58

Support pour la création de questions standardisées

But de la question:

.....

Thème:

.....

Affirmations

	important	pas important
positif	Aspects importants, fondamentaux	Aspects secondaires
négatif	Erreurs graves, opinions fausses fréquentes	Erreurs insignifiantes

Examen:

Domaine:

No de révision

But de la question:

Source, référence:

Réponse correcte:

Type de question:

Question:

Formulaire illustré en taille réduite

Auteur: Réviseur:

Contrôle de la formulation:

- bonne
- correction
- inappropriée

Pré-révision du contenu:

Rapport avec la formation

- en fait partie
- marginal
- aucun rapport avec la formation

Bien-fondé scientifique

- juste, certain
- controversé
- faux

Pertinence

- essentielle
- moyenne
- faible

Décisions de la révision:

Question acceptée

Pertinence: essentielle

à revoir par l'auteur

moyenne

refusée

faible

Informations générales sur la manière de procéder

Les candidats devraient être invités avant le début de l'épreuve à lire attentivement les indications suivantes, qui pourraient par exemple être imprimées au dos du cahier d'examen. Elles seront naturellement adaptées à chaque situation.

1. Commencez toujours par lire attentivement la question en entier, ainsi que toutes les réponses à choix.
2. Marquez tout d'abord vos réponses dans le cahier des questions. Pour les types de questions avec choix de la meilleure réponse, placez un cercle autour de la lettre correspondant à la réponse que vous pensez être la bonne. Pour les questions de type PickN, faites de même mais pour le nombre indiqué de bonnes réponses. Pour les questions de Type K', il faut par contre désigner chaque bonne réponse d'un signe (+) et chaque mauvaise réponse d'un signe (-).
3. Répondez à toutes les questions. Si vous n'êtes pas certain, choisissez la réponse qui vous semble le plus probablement correcte.
Vous recevrez un point pour chaque réponse (entièrement) juste et un demi-point si vous avez donné trois réponses correctes à une question de type K'. Vous pouvez également obtenir des demi-points en répondant correctement à une partie des réponses exactes pour les questions de type PickN. Les réponses fausses ne sont pas pénalisées. Une question sans réponse obtient 0 points, comme une réponse fausse.
4. Ne reportez vos réponses sur la feuille d'évaluation qu'après avoir répondu à toutes les questions de manière définitive et après avoir inscrit les réponses dans le cahier d'examen.
Pour les questions avec choix de la meilleure réponse, ne cochez qu'une case par question et, pour les questions de type PickN, que le nombre de cases exigé par la question correspondante.
Les réponses aux questions de type Kprime sont à reporter aux champs marqués «K». Remplissez une case pour chacune des 4 réponses partielles, soit une case (+) ou une case (-).
Les cases de réponses aux questions de type R et PickN sont à remplir au dos de la feuille d'évaluation.
5. La feuille d'évaluation ne doit être remplie qu'à l'aide du crayon fourni. Placez vos marques de façon claire et précise et évitez autant que possible l'usage de la gomme. Une marque floue, imprécise, ou des traits de crayons en dehors des champs prévus peuvent entraîner des erreurs à la lecture optique par l'ordinateur et donc des erreurs d'évaluation.
6. Signez le carnet et la feuille d'évaluation à l'emplacement indiqué. Vous confirmez ainsi avoir travaillé sans aucune aide illicite.

Instructions pour répondre aux différents types de questions

Le carnet d'examen doit contenir pour chaque groupe d'items d'un certain type des instructions détaillées sur la manière de répondre.

Type A

Choix d'une seule réponse

Ne marquez qu'une seule réponse en entourant d'un cercle la lettre correspondant à

- la seule réponse correcte ou la réponse qui est la plus appropriée en cas de formulation positive
- la seule exception, la seule réponse fautive ou la réponse la moins appropriée en cas de formulation négative (La négation est marquée en caractères gras.)

Type B

Attribution

Une série de cinq réponses à choix désignées par les lettres (A) à (E) est suivie d'un groupe de questions ou d'affirmations numérotées. Attribuez à chacune d'entre elles une des réponses, la seule correcte ou la plus appropriée, puis entourez d'un cercle la lettre correspondante. Une même réponse peut être correcte à plusieurs reprises.

Type R

Attribution élargie

Une liste comportant au maximum 26 propositions de réponses désignées par des lettres placées par ordre alphabétique est suivie d'un groupe de questions ou d'affirmations. Attribuez à chaque affirmation ou question **une des réponses de la liste**, la seule correcte ou la plus appropriée, puis inscrivez la lettre correspondante sous le numéro de la question. Une même réponse peut être correcte à plusieurs reprises.

Type PickN

Choix d'un nombre imposé de meilleures réponses possibles

Une liste comportant au maximum 26 propositions de réponses désignées par des lettres placées par ordre alphabétique est suivie d'un groupe de questions ou d'affirmations numérotées. Attribuez à chacune d'entre elles **le nombre de réponses de la liste qui est indiqué**, puis inscrivez les lettres correspondantes sous le numéro de la question. Une même réponse peut être correcte à plusieurs reprises.

Type Kprime / K'

Décision vrai/faux à quatre questions partielles

Une question ou une affirmation incomplète est suivie de quatre réponses ou fins de phrases. Décidez pour chacune d'elles si elle est vraie ou fautive et marquez un signe (+) ou un signe (-) en conséquence. Indépendamment du fait que la phrase soit formulée au singulier ou au pluriel, il est possible qu'il y ait 1, 2, 3, 4 ou même aucune réponse correcte.

Quatre bonnes appréciations donnent un point entier, trois un demi-point.

Type E

Relation de causalité

Deux affirmations sont reliées par l'expression «parce que». Il faut tout d'abord évaluer si chacune des deux affirmations est vraie ou fautive. Puis, si les deux sont justes, on doit décider si la relation de causalité exprimée par «parce que» est justifiée. Il y a donc cinq possibilités de réponse:

- | | | |
|-----|-------------|--|
| (A) | +parce que+ | Les deux affirmations sont correctes et la relation de causalité est justifiée. |
| (B) | +/- | Les deux affirmations sont correctes, mais la relation de causalité n'est pas justifiée. |
| (C) | +/- | La première affirmation est correcte et la seconde est fautive. |
| (D) | -/+ | La première affirmation est fautive et la seconde est correcte. |
| (E) | -/- | Les deux affirmations sont fautives. |

Glossaire des termes techniques relatifs aux examens

Analyse de l'item: appréciation des qualités psychométriques d'un item, en premier lieu de la ↗ difficulté et de la ↗ sélectivité de celui-ci. Pour les items des examens à choix multiple, on examine ensuite si chacune des réponses fausses détourne, comme on le souhaiterait, surtout les mauvais candidats de la réponse correcte (fonction des distracteurs/déviateurs).

Blueprint: catalogue pondéré du contenu de l'examen, selon lequel tous les examens sont composés. Il peut contenir une ou plusieurs dimensions.

Coefficient de corrélation: mesure statistique du rapport entre deux paramètres variables allant de -1 à +1. La valeur zéro signifie qu'il n'y a aucun rapport, la valeur 1 représente le meilleur rapport positif possible, les valeurs négatives signifient qu'il y a un rapport inverse.

„Cue“: mot anglais signifiant «signal, indice». On parle de "cue" quand la manière de poser la question ou la tâche à effectuer contient déjà un indice pour trouver la réponse correcte ou met celle-ci davantage en évidence. Ceux-ci devraient être évités dans l'intérêt de la ↗ fiabilité.

Déviateur ou distracteur: fausse réponse dans un item à choix multiple avec une seule réponse correcte possible. Les candidats aux connaissances lacunaires ne devraient pas pouvoir distinguer (par ex. grâce à des ↗ „cues“) la réponse correcte des réponses fausses. Par contre, il ne s'agit pas de piéger consciemment les candidats au moyen des déviateurs.

Difficulté de l'item: rapport du nombre de points obtenus par un groupe de candidats par rapport au nombre maximal de points pouvant être atteints. La difficulté est exprimée soit par la valeur percentile (P) ou par la valeur de probabilité (p)

Il ne faut pas oublier le fait qu'avec 5 réponses à choix, la chance de deviner ou de choisir au hasard se chiffre déjà à 20 % (P=20). Des items qui permettent une différenciation fiable des prestations ont une valeur P située entre 40 et 90.

Evaluation formative versus sommative: l'évaluation formative sert à orienter sur le niveau provisoirement atteint dans le processus d'apprentissage sans entraîner de sanctions pour le candidat. Par contre, l'évaluation sommative est une évaluation qui sanctionne au terme d'une étape de formation.

Fiabilité (en allemand Reliabilität, en anglais reliability): critère de qualité d'un examen qui cherche à savoir avec quelle exactitude un paramètre est mesuré, indépendamment du fait s'il s'agit du paramètre qu'on avait l'intention de mesurer (cf validité). Le coefficient de fiabilité varie entre 0 et la valeur maximale 1. Les erreurs pouvant biaiser la fiabilité sont par ex. le manque d'↗ objectivité, les facteurs influençant la probabilité de deviner la réponse correcte, le choix trop restreint et donc non représentatif des items et les hasards. La fiabilité de l'instrument d'examen est aujourd'hui le paramètre primordial, et elle est la plupart du temps exprimée par le coefficient de Cronbach alpha. Ce coefficient renseigne sur la force de l'influence sur les résultats du choix spécifique des items, respectivement à quel point un autre examen avec le même nombre d'items choisis selon les mêmes critères dans les mêmes domaines permettrait de trier les candidats de manière identique que l'examen donné.

Pour des résultats d'examen aux conséquences cruciales pour les personnes concernées, la fiabilité devrait atteindre au moins 0.8; une valeur de 0.9 est considérée comme optimale.

Item: élément. Vu dans le cadre d'un examen, cet élément peut représenter une question, une tâche, une observation ou une appréciation unique.

Objectivité: dans le cadre d'examens, l'objectivité signifie l'indépendance des résultats d'examen par rapport aux examinateurs. On distingue l'objectivité de l'exécution, de l'évaluation et de l'interprétation. Elle est calculée le plus souvent par la concordance statistique entre les différents examinateurs. On utilise le terme d'„interrater-reliability“ en rapport avec la ↗ fiabilité.

Sélectivité de l'item (r_{it}): mesure de sa capacité à différencier dans un examen les candidats fournissant une bonne performance de ceux dont le résultat est faible. Elle se calcule comme ↗ coefficient de corrélation (r) entre le nombre de points atteints pour un item et le nombre total de points atteints pour l'examen entier sans l'item en question.

Des items qui permettent une différenciation fiable des prestations ont une valeur de sélectivité clairement positive, si possible ≥ 20 , mais certainement ≥ 10 . Des items avec une sélectivité avoisinant 0 ne contribuent pas à faire cette différenciation, ceux avec une sélectivité négative vont à l'encontre de ce but et devraient, d'un point de vue psychométrique, être éliminés.

Validité: critère de qualité d'un examen qui vérifie si le procédé donné mesure réellement ce qu'on avait l'intention de mesurer. A la fin d'une formation, l'examen devrait tester si les compétences nécessaires à l'exercice de la profession sont acquises.

On distingue les aspects suivants:

- **validité du contenu („content validity“)**: demande dans quelle mesure le contenu de l'examen est représentatif comme échantillon du domaine de compétences/de comportements (ou de la branche) à mesurer: a) sur le plan de chaque item; b) sur le plan de l'examen total (↗ blueprint). Cette appréciation devrait se faire par des experts.
- **validité critériée („criterion related validity“)**: mesure la concordance des résultats de l'examen avec les performances en dehors de la situation d'examen par exemple dans la pratique quotidienne. On essaie surtout de la clarifier à l'aide d'études de corrélation.
- **validité de concept („construct validity“)**: s'intéresse à savoir si les hypothèses issues d'une théorie sur le concept qu'on prétend mesurer avec l'examen (par ex. la capacité à résoudre des problèmes) peuvent être appuyées par des constatations issues des résultats de l'examen.
- **validité apparente („face validity“)**: représente comme son nom l'indique une apparence et par conséquent n'est pas un critère de qualité comme les 3 validités précédentes. Un examen qui a une validité apparente donne l'impression à un observateur de mesurer ce que l'examen est censé mesurer. Ce but peut valoir la peine d'être poursuivi, parce qu'il permet une meilleure acceptation de l'examen par les candidats et par les décideurs.