

Handleiding bij het maken van multiple-choice vragen

Inhoud

Inleiding.....	2
Onderdelen van een Multiple-choice vraag ⁹	3
1 Alle opties zijn bij voorkeur op hetzelfde aspect gericht.....	4
2 De opties zijn bij voorkeur even lang.....	4
3 De opties zijn alle in dezelfde mate genuanceerd.....	4
4 De opties zijn alle in dezelfde richting (alle ontkennend of bevestigend).....	5
5 Test bij voorkeur maar één aspect per optie.....	5
6 Formuleer helder en leesbaar.....	5
7 De opties moeten het hele mogelijke antwoordscale omvatten.....	6
8 Opties moeten elkaar uitsluiten.....	6
9 Een antwoord moet verdedigbaar correct zijn, de andere verdedigbaar incorrect.....	7
10 Gebruik geen ‘verzamelopties’.....	7
11 Vermijd grammaticale aansluitingsproblemen.....	8
12 Geef geen logische cues.....	8
13 Gebruik geen absolute of te open opties.....	8
14 Vermijd semi-kwantitatieve termen.....	9
15 Plaats de opties in een logische of alfabetische volgorde.....	9
16 Gebruik geen ‘U vragen’.....	9
17 Vermijd gecompliceerde vormen van MC-vragen.....	9
18 Bied nooit ‘stellingvragen’ aan.....	10
Ten slotte.....	11
Literatuur.....	12
Addendum: Checklist ten behoeve van het schrijven van meerkeuzevragen.....	13
Addendum: Voorbeelden toetsvragen.....	14
Goede toetsvragen.....	14
Toetsvragen die bijna goed zijn.....	15
Ongeschikte toetsvragen.....	15

Inleiding

Welke eigenschappen hebben meerkeuze(MC)vragen? Hieronder een opsomming van wat hierover is geschreven¹⁻³.

- *Meet een open vraag iets anders dan een MC vraag?*⁴
Nee, de vorm van een vraag bepaalt niet wat een vraag meet. De inhoud is van eminent belang. Dus **wat** een vraag vraagt is van belang en niet hoe het antwoord wordt gegeven. Dit is een robuust onderzoeksgegeven. Wel is het zo dat men een MC vraag kan beantwoorden door herkenning van het juiste antwoord (cueing effect), maar het blijkt dat dat niet een fundamenteel verschil in kennis is. Daarnaast is het natuurlijk zo dat sommige inhouds niet passen binnen bepaalde vraagvormen (spontaan aan een diagnose denken is bijvoorbeeld niet iets dat in een MC vraag gevraagd kan worden).
- *Bereiden studenten zich anders voor op MC vragen dan op open vragen?*⁵
De beperkte onderzoeken hiernaar kunnen niet echt een verschil in daadwerkelijke voorbereiding vinden, maar in de vele onderzoeken die studenten bevragen wordt wel gevonden dat studenten zeggen zich anders voor te bereiden.
- *Zijn MC vragen betrouwbaarder dan open vragen?*⁶
Vaak wordt gedacht dat bij het nakijken van open vragen subjectiviteit geïntroduceerd wordt. Toch hoeft dit niet zo te zijn. Een uitgewerkte antwoordsleutel, een verdeling van nakijkers over de vragen (docent A kijkt van iedere student vraag 1 na, docent B van iedere student vraag 2, etc.) is een eenvoudige manier om de toets betrouwbaarder na te kijken dan een verdeling over studenten (docent A kijkt alles na van studenten 1 - 10, docent B alles van studenten 11 - 20, etc.).
Regelmatig overleg tussen nakijkers en tweede nakijkers bij twijfelgevallen zijn maatregelen die over het algemeen de invloed van subjectiviteit op de betrouwbaarheid voldoende verminderen.
Wanneer men (zoals bij de voortgangstoets) een heel breed domein wil bevragen, kan men meer MC-vragen stellen per uur toetstijd dan open vragen en daardoor wordt de domeindekking beter .
- *Het gaat niet alleen om betrouwbaarheid en validiteit maar ook om kosten en bruikbaarheid*^{7,8}.
Het is wel leuk om in theorie alles met open vragen of essays te willen doen, maar bij grote aantallen studenten wordt dat logistiek moeilijk. Bij de voortgangstoets bijvoorbeeld spreken we over 200 vragen vier keer per jaar bij 8.000 - 10.000 studenten per toetsafname. Dat is niet te doen met open vragen of essays. Het grote nadeel van een voorgeschreven vraagvorm, in dit geval dus MC-vragen, is dat dit ook het gedrag van de docent beïnvloedt. Het kan makkelijk leiden tot het vermijden van onderwerpen die lastig in MC-vragen te bevragen zijn of tot een soort fatalistische overtuiging dat met MC-vragen alleen triviale feiten bevroegd kunnen worden. Deze handleiding beoogt onder meer hieraan tegemoet te komen. Voor wat betreft de validiteit van vragen geeft onderzoek en literatuur aan dat gesloten vragen (vraag met vraagteken) het meest valide zijn.

Onderdelen van een Multiple-choice vraag⁹

Een multiple choice vraag bestaat uit een stam waarin de context beschreven wordt waarin de vraag gesteld wordt. De eigenlijke vraag wordt ook wel *lead-in* genoemd. De vraag is zo opgesteld dat de student in staat zou moeten zijn het antwoord te formuleren zonder naar de alternatieven te hebben gekeken. De opties bestaan uit een juist alternatief (het antwoord) en foute alternatieven, de afleiders (distractors). De afleiders worden zodanig geformuleerd dat de set alternatieven volkomen gelijkwaardig is.

Stam:

De heer van Dam is 35 jaar oud. Hij heeft een bacteriële bronchopneumonie. Hij is niet recent opgenomen geweest in een ziekenhuis of verpleeghuis, ook is hij daar niet werkzaam. Meneer van Dam is verder niet immuungecompromitteerd.

Lead-in:

Wat is de meest waarschijnlijke verwekker van zijn bronchopneumonie?

Opties:

- A Hemophilus influenzae;
- B Klebsiella pneumoniae;
- C Pneumocystis carinii;
- D Staphylococcus aureus;
- E Streptococcus pneumoniae.

antwoord: B

Hoewel het heel vaak voorkomt dat in een MC toets alle vragen hetzelfde aantal opties hebben (meestal 4 of 5) is dat niet noodzakelijk. Het is onderzocht en er blijkt geen enkele psychometrische reden om het aantal opties vast op 4 of 5 te houden. Er is wel een inhoudelijk argument om het aantal opties per vraag af te laten hangen van het aantal realistische opties¹⁰. Het minimale aantal opties is 2. Dan dient er wel rekening mee te worden gehouden dat ten behoeve van de betrouwbaarheid van het tentamen een groter aantal vragen gesteld dient te worden.

Bij het schrijven van een MC-vraag helpt het om de vraag allereerst als een *kort antwoord open-vraag* te benaderen. Dat wil zeggen dat de docent allereerst vaststelt wat hij precies wil vragen. Vervolgens verzamelt hij de afleiders die net als het juiste antwoord op één bepaald kennisaspect zijn gericht.

Hieronder volgen de belangrijkste regels voor het maken van multiple choice vragen^{11, 12}.

1 Alle opties zijn bij voorkeur op hetzelfde aspect gericht.

Het beste is dit te testen door de zogenaamde afdekproef. Als je de opties afdekt, is de vraag dan nog te beantwoorden? Bij de onderstaande vraag is dit niet het geval:

Amsterdam is: A een grote stad; B gelegen aan het IJ; C de hoofdstad van Nederland.
--

Als men nu de opties niet zou zien, dan kan men de vraag niet beantwoorden. Dit is geen triviale regel (vandaar dat hij ook als eerste genoemd wordt). Uit onderzoek¹³ blijkt dat ook bij MC vragen een behoorlijk deel van de studenten de vraag leest, dan gaat nadenken over het juiste antwoord en dan pas naar de opties kijkt. Dit is een vorm van redeneren, *forward reasoning*, die we graag bij studenten willen stimuleren. Het gebrek van de vraag hierboven ligt in de formulering "is ...". De lead-in moet een echte vraag zijn, gevolgd door een vraagteken.

Als men vragen stelt zoals bovenstaand voorbeeld wordt forward reasoning ontmoedigd. Dit geldt natuurlijk ook voor lead-ins als:

- welke van onderstaande is juist
 - wat geldt voor/wat geldt niet voor
 -is:
- etc.

Al deze vormen van lead-ins mogen dus nooit worden gebruikt!

2 De opties zijn bij voorkeur even lang

Idealiter is een vraag een perfecte voorspeller voor aanwezige kennis, dus diegenen die het weten moeten het juiste antwoord kunnen geven en diegenen die de kennis niet hebben moeten niet het juiste antwoord kunnen geven. Voor het gemak noemen wij de situatie waarbij iemand **zonder** kennis toch het juiste antwoord geeft fout-positieve response en het omgekeerde fout-negatieve response. Het mag duidelijk zijn dat beide foutenbronnen een verstorende invloed hebben op de validiteit van de toets, zeker als men de toets zou willen vergelijken met een diagnosticum voor aanwezige competentie. Studenten weten wel een aantal trucs om een vraag toch goed te kunnen beantwoorden. De belangrijkste hiervan is dat de langste optie vaak de juiste is. Er zijn vaak meer woorden nodig om een optie juist te laten zijn dan om een optie onjuist te laten zijn.

Wat is de beste behandeling voor een pneumonie?

- | |
|---|
| A Antibiotica.
B Aciclovir.
C Antimycotica.
D Dit moet gekozen worden aan de hand van de specifieke oorzaak. |
|---|

Dit is natuurlijk een wat absurdistisch voorbeeld, maar het is een fout die heel vaak gemaakt wordt bij het schrijven van MC-vragen en die leidt tot fout-positieve response.

3 De opties zijn alle in dezelfde mate genuanceerd

Vaak zit er een nuanceverschil in de opties. Ook dit is logisch: de werkelijkheid is vaak genuanceerder dan men op papier kort kan verwoorden, daarom zal de meest genuanceerde optie ook wel de juiste zijn. Voor foute opties formuleert de vragenmaker vaak minder genuanceerd.

Wat is de meest aangewezen behandeling bij chronisch benigne lage rugpijn?

- A Tramadol voorschrijven
- B Fysiotherapie/Mensendieck/Cesar
- C Discusprothese aanbrengen
- D Multidisciplinaire aanpak

D is weliswaar niet de langste optie, maar zij is zeker de genuanceerde optie. Het lijkt waarschijnlijk dat de auteur dit als juiste antwoord bedoeld heeft. In zo'n geval zal er sprake zijn van fout-positieve response.

4 De opties zijn alle in dezelfde richting (alle ontkennend of bevestigend)

Het is erg verwarrend als er opties zijn die aangeven wat *wel* van toepassing is in de vraag terwijl er ook opties zijn die op het *niet* van toepassing zijn gericht zijn. Dat veroorzaakt leesfouten, dus *fout-negatieve response*. Vragen gesteld in een positieve richting hebben een betere response dan vragen die negatief gesteld zijn en verdienen in toetsen de voorkeur.

5 Test bij voorkeur maar één aspect per optie

Twee-in-één opties leiden tot onduidelijkheden en tot de convergentiestrategie. Waar mogelijk dus vermijden. Alleen als er een inhoudelijke voorwaarde voor is dan zal het moeten, maar dan moeten wel alle theoretische mogelijkheden gedekt zijn.

Als het al nodig is om twee dingen per optie te bevragen, zorg er dan voor dat alle mogelijkheden gedekt zijn (tenzij er dan een echte onzinoptie bijkomt). Hier volgt een voorbeeld:

Niet alleen repetitive nerve stimulation (RNS) en single fibre EMG (SFEMG), maar ook hoge titers van antilichamen tegen acetylcholine receptoren (AbAChR) kunnen gebruikt worden om de diagnose myasthenia gravis te stellen. Welke van de onderstaande opties is correct betreffende de sensitiviteit van deze testen?

- A AbAChR hoger dan SFEMG en hoger dan RNS
- B AbAChR lager dan SFEMG en lager dan RNS
- C AbAChR lager dan SFEMG en hoger dan RNS
- D AbAChR hoger aan SFEMG en lager dan RN

Aan deze vraag valt op dat de auteur waarschijnlijk deze vorm heeft gekozen om zonder moeite aan precies 4 alternatieven te komen. Dit is een af te raden strategie! Het getal 4 is verre van heilig.

6 Formuleer helder en leesbaar

Een moeilijk leesbare tekst is een storend element in een **kennistoets**. Daarmee is niet gezegd dat een academisch geschoold iemand niet goed zou moeten kunnen lezen, maar misschien is het dan beter om hiervoor andere en betere instrumenten in het toetsprogramma op te nemen. Probeer dus een stam helder te schrijven, de zinnen in een logische volgorde te zetten en onnodig gecompliceerde formuleringen te vermijden. De onderstaande vraag is een absurd voorbeeld van hoe geprobeerd is om door middel van ingewikkelde constructies de vraag 'moeilijk' te maken. Na ontrafelen van de zin is de vraag niet meer moeilijk

Het kan niet uitgesloten worden dat bepaalde symptomen niet gevonden worden zonder enige vorm van pathologie bij een patiënt die een meningitis purulenta heeft, wanneer deze patiënt niet deel uitmaakt van de normale volwassen populatie. Waarvoor geldt dit bijvoorbeeld?

- A Leukocytose van het spinaalvocht
- B Nekstijfheid
- C Ontstekingsreacties in de meningen

Verder heeft deze vraag ook een gebrekkige lead-in. Deze behoort te luiden:

Voor welke bevinding geldt dit?

Bovendien is C niet geheel gelijkwaardig aan A en B!

7 De opties moeten het hele mogelijke antwoordscale omvatten

Het is natuurlijk jammer als er theoretische opties zijn die niet gegeven worden. Waar mogelijk moet met de opties het hele scala omvat worden. Een wat overdreven voorbeeld is:

De sensitiviteit van een standaard AP thoraxfoto voor het detecteren van een plaveiselcelcarcinoom is:

- A groter dan 95%
- B kleiner dan 90%

De optie tussen 90 en 95% zit hier niet in. Daarmee wordt de keuze van de student gestuurd. Het omgekeerde treedt op als alle mogelijkheden al in een deel van de opties genoemd zijn:

Wat is in de meeste gevallen het effect van propranolol?

- A een verlaging van de bloeddruk
- B een verhoging van de bloeddruk
- C niet van invloed op de bloeddruk
- D vrij laat optredend
- E gepaard gaand met weinig bijwerkingen

Nu is het duidelijk dat de opties D en E niet overwogen hoeven te worden. A, B en C dekken het hele domein, een stof verhoogt of verlaagt de bloeddruk of heeft er geen effect op. Het kan dus niet zo zijn dat een optie A en optie B en optie C fout zijn. Bovendien zijn de opties A, B en C van een geheel andere aard dan opties D en E. Deze laatste twee lijken er achteraf te zijn bij gezocht. Zie onder 1: de opties dienen op hetzelfde aspect gericht te zijn, zij behoren volkomen gelijkwaardig te zijn.

8 Opties moeten elkaar uitsluiten

Als er overlap is tussen de opties krijgt de student een krachtige cue om uit te vinden wat het juiste antwoord is.

De overall 5-jaarsoverleving bij patiënten bij wie een gemetastaseerd kleincellig longcarcinoom gediagnosticeerd is ligt:

- A onder de 10%
- B tussen de 10 - 30%
- C tussen de 20 - 40%
- D tussen de 30 - 50%
- E tussen de 40 - 60%

Er zit zoveel overlap tussen de opties B tot en met E dat heel duidelijk wordt dat het juiste antwoord niet kan liggen tussen 20 en 30% (overlap tussen B en C), 30 en 40% (overlap tussen C en D), etc. Nu wordt de keuze eigenlijk beperkt tot A en E.

9 Een antwoord moet verdedigbaar correct zijn, de andere verdedigbaar incorrect

Dit klinkt als een open deur maar dat is het niet. Vaak blijkt dat een vraag zo gesteld is dat de juiste optie niet helemaal correct is of dat afleiders ook goed kunnen zijn.

Een 40-jarige vrouw heeft sinds kort last van maagpijn, met name na de maaltijd. Onderzoek leidt tot de diagnose ulcus duodeni. Welke van de onderstaande geneesmiddelsgroepen is het meest aangewezen voor de start van medicamenteuze behandeling?

- A. Een antacidum.
- B. Een propulsivum.
- C. Een H2-receptorantagonist.
- D. Een protonpompremmer.

Deze vraag is bij een nabespreking komen te vervallen omdat bleek dat zowel C als D als antwoord te verdedigen waren. Dit is een inhoudelijke kwestie.

Soms zijn ook meerdere antwoorden te verdedigen omdat de vraag niet scherp genoeg gesteld is:

De ovulatie treedt op na een LH (Luteïniserende Hormoon)-piek. Er ligt een bepaalde periode tussen de LH-piek en het moment van ovulatie. Hoe lang is deze periode?

- A. 18 uur;
- B. 36 uur;
- C. 54 uur;
- D. 72 uur.

Hier is niet duidelijk genoeg aangegeven vanaf welk moment van de LH-piek gemeten is.

10 Gebruik geen ‘verzamelopties’

Niet alles met betrekking tot itemconstructieregels is even sterk onderzocht, maar het is wel evident uit onderzoek dat verzamelopties (alle bovenstaande, geen van bovenstaande) een storende werking hebben op de zuiverheid van de competentiemeting. Bij de verzameloptie “geen van bovenstaande” moet men zich afvragen wat de auteur van de vraag eist van de kennis bij de student. Waarom worden opties geformuleerd die geen van allen juist zijn? Wel is het mogelijk om uit een lijst te laten kiezen welk antwoord (methode, medicament enzovoorts) pertinent ONJUIST is. De lead-in vraagt dan ook expliciet hiernaar.

Bij de verzameloptie “alle bovenstaande” is het storende effect gemakkelijker verklaarbaar. Stel dat er vijf opties zijn, vier inhoudelijke en de vijfde als verzameloptie, dan is het voor iedere student die er twee weet al duidelijk dat de verzameloptie de juiste is. Dat wil niet zeggen dat er in toetsing geen plaats zou zijn voor vragen waarin meerdere antwoorden goed kunnen zijn (zogenaamde multiple true-false items), maar indien een gewone MC vraag bedoeld is, dienen verzamelopties te worden vermeden.

Welke van onderstaande aandoeningen kunnen ten gevolge van cryptorchisme op latere leeftijd voorkomen?

- a) Torsio testis
- b) Onvruchtbaarheid
- c) Hernia inguinalis
- d) Maligne ontaarding
- e) **Alle alternatieven zijn juist**

11 Vermijd grammaticale aansluitingsproblemen

Studenten zullen alle informatie gebruiken om het goede antwoord te raden als ze het niet weten. Grammaticale aansluitingsproblemen zijn dan een eenvoudige hint voor studenten. Hetzelfde geldt voor de logische cues die hieronder genoemd worden.

De stabiliteit van welke band wordt door middel van de Lachman's test onderzocht?

- a) De mediale kniebanden
- b) De laterale kniebanden
- c) De voorste kruisband**
- d) De achterste kruisband

Het enkelvoud van het woord band suggereert dat alleen c of d goed kan zijn.

12 Geef geen logische cues

De volgende vraag is een voorbeeld van inhoudelijke aansluitingsproblemen.

Een 60-jarige alcoholist wordt door de politie naar de Eerste Hulp gebracht. Hij heeft continu epileptische aanvallen die niet stoppen. De arts controleert eerst of de ademweg vrij is.

Welke intraveneuze injectie is de volgende stap in de work-up?

- A fenytoïne
- B glucose met vitamine B1
- C CT-scan van het hoofd
- D onderzoek van de liquor
- E diazepam

Het is voor de student al snel duidelijk dat opties C en D niet passen bij de "intraveneuze injectie". Het feit dat niet alle afleiders gelijkwaardig zijn geeft het probleem van deze vraag direct aan.

13 Gebruik geen absolute of te open opties

Opties met 'kan', 'is mogelijk' en dergelijke zijn zo open gesteld dat het vaak moeilijk te verdedigen dat ze echt fout zijn (zie ook regel 9). Anderzijds zijn opties met 'altijd', 'nooit', 'uitgesloten', en dergelijke zo absoluut gesteld dat ze zeer waarschijnlijk niet correct zijn.

Patiënten met diabetes mellitus hebben nooit hart- en vaatziekten
juist / onjuist

Patiënten met diabetes mellitus zijn altijd ouder dan 25 jaar.
juist / onjuist

Patiënten met diabetes mellitus kunnen last hebben van slecht genezende wonden.
juist / onjuist

Patiënten met diabetes mellitus moeten met insuline worden behandeld.
juist / onjuist

Bovenstaande juist/onjuist-vragen zijn alle vier gemakkelijk te beantwoorden zonder enige geneeskundige kennis. De woorden "nooit", "altijd" en "moeten" zijn zo absoluut dat zij niet juist kunnen zijn. Tegelijkertijd is "kunnen last hebben" zo algemeen gesteld dat het niet onjuist kan zijn.

14 Vermijd semi-kwantitatieve termen

Hoewel onze leerboeken vol staan met vage semi-kwantitatieve termen (vaak, soms, niet zelden, frequent, etc.) dienen ze te worden vermeden in toetsvragen. Het is namelijk niet duidelijk hoe vaak 'vaak' is, hoe zelden 'zelden', etc. Uit onderzoek blijkt dat er een enorme spreiding is in wat mensen onder deze termen verstaan als je ze vraagt om ze in een procentrange uit te drukken. Veelal wordt gezocht naar een omvorming tot percentagevragen. Dat is wel exacter, maar tegelijk ontstaat dan de vraag hoe exact de student iets moet weten.

15 Plaats de opties in een logische of alfabetische volgorde

Door de opties in een logische volgorde te zetten (bijvoorbeeld toenemende mate van ernst of toenemende mate van invasiviteit bij diagnostiek) of door de opties te alfabetiseren wordt voorkomen dat de student aan de gekozen volgorde meent te zien wat het juiste alternatief is.

16 Gebruik geen 'U vragen'

Bij vragen gesteld in de "U-vorm" wordt gevraagd naar een persoonlijke mening. De antwoorden op deze vragen worden door sommigen als altijd juist beoordeeld.

Een jong meisje met een negroïde huidpigmentatie (huidtype VI) heeft gaatjes in haar oorlellen laten prikken. Na 2 maanden zijn er stevige ronde, huidkleurige tumoren ontstaan met een diameter van ca. 1 cm. aan beide oorlellen. Wat is uw diagnose?

- A. Atheroomcyste;
- B. Granuloma teleangiëctaticum;
- C. Keloïd;
- D. Sarcoïdose.

Antwoord: C

Dat sommigen bezwaar hebben tegen de vraag "Wat is uw diagnose?" lijkt wat overdreven, maar het is zeker dat de vraag "Wat is de juiste diagnose?" zuiverder is.

17 Vermijd gecompliceerde vormen van MC-vragen

Er is een tijd gedacht dat vragen waarbij op gecompliceerde wijze informatie wordt aangeboden en gecombineerd moet worden meer inzicht toetsen. Dat bleek echter onwaar te zijn¹⁴. Essentieel is het onderscheid tussen vragen met een casus of probleem, die gericht zijn op beslissingen, en vragen die rechttoe rechtaan een feit vragen (dus zonder probleem of casus). Beide typen kunnen relevant zijn, maar de denkstappen bij de casusvorm zijn veel meer gericht op het afwegen van waarschijnlijkheden, vragen zonder casus leiden tot denkstappen die gebaseerd zijn op een serie van ja/nee-beslissingen. Gecompliceerde vormen leiden alleen maar tot onnodige complicatie en niet tot betere competentiemeting. Een ander aspect is dat ze vaak ook nog inhoudelijke of logische cues geven over het juiste antwoord:

De belangrijkste symptomen die geassocieerd zijn met cardiovasculaire aandoeningen:

- 1 zijn pijn op de borst, dyspnoe, palpaties
- 2 verschijnen of nemen toe bij inspanning
- 3 zijn moeheid, duizeligheid en flauwvallen
- 4 verschijnen in rust

- A (1), (2) en (3) zijn correct
- B (1) en (3) zijn correct
- C 2) en (4) zijn correct
- D alleen (4) is correct
- E alle zijn correct

De lezer kan menen dat deze vraag werd verzonnen en om didactische redenen in dit overzicht is opgenomen. Dat is niet het geval. Het betreft een vraag die door een docent met serieuze bedoelingen werd vervaardigd. Aan zijn behoefte aan een gecompliceerd toetsitem heeft hij zonder meer voldaan. Maar de vraag zondigt tegen veel van de hierboven gegeven regels. Het belangrijkste bezwaar is dat hier sprake is van vier stellingen, waarvan de juistheid moet worden beoordeeld. De stellingen betreffen niet hetzelfde aspect, dat is: zijn ongelijkwaardig, terwijl bovendien 2 en 4 strijdig met elkaar zijn.

Dit brengt ons op een veel voorkomende fout bij de constructie van meerkeuze- vragen: de zogenaamde stellingvraag.

18 Bied nooit 'stellingvragen' aan

Vaak komt men vragen tegen van het volgende type:

- I. Allochtonen die een tolk meebrengen, krijgen vaak een onjuiste vertaling te horen van de woorden van de huisarts.
 - II. Juist bij allochtone patiënten is het van belang om op non-verbaal gegeven informatie te letten.
-
- a. I en II zijn beide juist.
 - b. I en II zijn beide onjuist.
 - c. I is juist, II is onjuist.
 - d. I is onjuist, II is juist.

Waarom staan deze twee stellingen in één vraag? Sommige auteurs menen helaas nog steeds dat een meerkeuzevraag een 4-keuze-vraag is of behoort te zijn. En met bovenstaand format komt men moeiteloos tot vier alternatieven. Tegen dit misverstand moet opnieuw worden gewaarschuwd. Het aantal alternatieven kan maximaal 6 (voor ICLON) of 5 (voor de Voortgangstoets) zijn.

Houden de twee stellingen verband met elkaar, of wil de vragensteller dat suggereren? Wat wil men dat de student weet? Als hij de ene stelling goed beoordeelt, maar de andere niet, krijgt hij niettemin geen punten. Als men wil dat de student beide stellingen kan beoordelen, dan is de enige oplossing het aanbieden van twee juist/onjuist-vragen. Onderzoek wijst uit dat deze verandering de betrouwbaarheid van het tentamen doet toenemen.

Hierbij nog een kort overzicht van de regels voor het maken van een goede MC-vraag:

- 1 Alle opties zijn bij voorkeur op hetzelfde aspect gericht.
- 2 De opties zijn bij voorkeur even lang
- 3 De opties zijn alle in dezelfde mate genuanceerd
- 4 De opties zijn alle in dezelfde richting (alle ontkennend of bevestigend)
- 5 Test bij voorkeur maar één aspect per optie
- 6 Formuleer helder en leesbaar, maak met name een heldere stam
- 7 De opties moeten het hele mogelijke antwoordscale omvatten
- 8 Opties moeten elkaar uitsluiten
- 9 Een antwoord moet verdedigbaar correct zijn, de andere verdedigbaar incorrect
- 10 Gebruik geen "verzamelopties"
- 11 Let op grammaticale aansluitingsproblemen
- 12 Geef geen logische cues
- 13 Gebruik geen absolute of te open opties
- 14 Vermijd semi-kwantitatieve termen
- 15 Plaats de opties in een logische of alfabetische volgorde
- 16 Vermijd het gebruik van U vragen.
- 17 Vermijd gecompliceerde vormen van MC-vragen
- 18 Biedt nooit stellingvragen aan

Ten slotte

Deze korte handleiding is niet uitputtend. Het belangrijkste doel ervan is te laten zien dat goede vragen maken niet ondoenbaar is en dat veel gemaakte fouten gemakkelijk zijn te vermijden. De belangrijkste boodschap is dat idealiter een vraag tot stand komt door de samenwerking tussen de inhoudelijke expert en iemand met ervaring en kennis op het gebied van itemconstructie. Het is ook altijd een afweging, soms moet inhoudelijk iets worden toegegeven en soms moet in de formulering iets worden toegegeven. Als beiden het er maar over eens zijn dat vorm en inhoud van de vraag optimaal zijn. De 18 regels kunnen samen met onderstaande checklist leiden tot een goede toets!

januari 2011

Literatuur

1. Van Berkel, Bax: Toetsen in het Hoger Onderwijs (2002), p.119 – 135.
2. International Handbook of Research in Medical Education, Volume 7: ISBN 1-4020-0466-4.
3. Norman GR, Smith EKM, Powles ACP (2006): Factors underlying performance on written tests of knowledge, Medical Education. 21:297-304.
4. Smith SM (2005): Learning approaches: Examination type, discipline of study, and gender. Educational Psychology: 25(1):43-53
5. Tang KC (1992): Perception of task demand, strategy attributions and student learning. Research and Development in Higher Education. 15:474-481
6. Downing, Haladyna (1996): Validity and Reliability in MC-Questions and Open-Source Questions. Research and Development in Higher Education.
7. Knapp & Knapp (1995):
8. Scherpbier AJJA, Van der Vleuten CPM, Rethans JJ, Van der SteegAFW: Advances in Medical Education. 1997:2-6.
9. Case M, Swanson DB (2001): Constructing Written test Questions for the basic and clinical science. NBMA, 3th edition.
10. Lord (1977)
11. Haladyna TM, Downing SM (1988a) A taxonomy of multiple-choice item- writing rules. Applied Measurement in Education, 1, 37-50
12. Haladyna TM, Downing SM (1988b) The validity of a taxonomy of multiple-choice item-writing rules. Applied Measurement in Education, 1, 51-78
13. Frederikson N (1984). The real test bias: influence of testing and teaching and learning. America Psychologist 39:193-202.
14. Case SM, Swanson DB, Becker DF: Verbosity, window dressing, and red herrings: do they make a better test item? Academic Medicine, 1996;71:528-530.

Addendum: Checklist ten behoeve van het schrijven van meerkeuzevragen

Een meerkeuzevraag bestaat uit een stam (de vraag) en enkele antwoordalternatieven waaruit het juiste antwoord gekozen moet worden.

Let op: als meer antwoorden goed zijn, maar slechts één antwoord het best, dan moet er een duidelijke instructie hieromtrent op het voorblad staan.

Checklist

1. Stel bij iedere vraag de volgende vragen:
 - * Wil ik dit echt van de studenten weten?
 - * Op welk onderwijsdoel heeft deze vraag betrekking?
2. Bevat de stam een duidelijke vraag (waarom, wanneer, wat, hoe enz.)?
3. Bevat de vraag één probleem?
4. Sluiten de alternatieven aan op dit ene probleem? Komen in de alternatieven niet meerdere kwesties aan de orde?
5. Is de vraag kort en zakelijk geformuleerd?
6. Zijn ontkenningen als niet vermeden?
7. Als een ontkenning als niet absoluut moet worden gebruikt, is het woord niet dan onderstreept?
8. Bevat de vraag geen dubbele ontkenning?
9. Is slechts één alternatief goed of duidelijk het beste?
10. Is de vraag met behulp van de bestudeerde leerstof te beantwoorden (en niet op basis van kennis van de wereld e.d.)?
11. Bevat de vraag geen woorden als: altijd, nooit, meestal, zeker?
12. Zijn alle alternatieven ongeveer even lang?
13. Bevatten de alternatieven geen overlappingen?
14. Zijn er geen alternatieven die vooraf te elimineren zijn?
15. Sluiten de alternatieven grammaticaal aan bij de stam?
16. Is de gebruikte terminologie consequent?
17. Zijn er geen afkortingen gebruikt?
18. Wordt een saillant woord uit de stam niet herhaald in één van de alternatieven?
19. Lijken de alternatieven inhoudelijk niet te veel op elkaar?
20. Is één alternatief niet duidelijk anders dan de andere?
21. Zijn alle alternatieven even aantrekkelijk?
22. Bevat de vraag geen valkuilen of listigheden?
23. Zijn de alternatieven in oplopende/alfabetische volgorde geplaatst?
24. Is bij een vraag naar een opinie de context/het artikel/het boek waarin de studenten die zijn tegengekomen gegeven?
25. Is bij een vraag met een citaat de bron van dat citaat gegeven?
26. Geeft de formulering van de stam en / of een of meerdere alternatieven van de ene vraag geen aanwijzing voor het goede antwoord op een andere vraag?
27. Is het gebruik van een alternatief in de trant van 'geen (of alle) van voorgaande alternatieven' vermeden?

Addendum: Voorbeelden toetsvragen

NB. met toestemming van vraagmakers Nederlandse nascholingen.

Goede toetsvragen

1. In 1974 startte in Nederland het hielprikprogramma. Wat was de eerste ziekte die in dit programma werd opgenomen?

- A. Cystic fibrosis (CF)
- B. Spinale musculaire atrofie (SMA)
- C. Phenylketonurie (PKU)

Dit is een goede vraag gezien antwoorden even lang zijn en hetzelfde aspect betreffen. Bovendien zijn deze antwoorden verdedigbaar correct en incorrect. Een verbetering is nog wel mogelijk in de stam van de vraag, de klinische relevantie kan hierbij nog aangegeven worden.

2. Wanneer een ziekte wordt opgenomen in de hielprik, moet er aan een aantal criteria voldaan worden (volgens Wilson and Junger). Wat is zo'n criterium waaraan voldaan moet worden om een ziekte in de hielprik op te nemen?

- A. Er moet een herkenbaar symptomatisch stadium van de ziekte zijn.
- B. Het natuurlijke verloop van de op te sporen ziekte moet bekend zijn.
- C. Er moeten minimaal 100 patiënten bekend zijn met de ziekte.

Dit is een goede vraag gezien antwoorden even lang zijn en hetzelfde aspect betreffen. Bovendien zijn de antwoorden in dezelfde mate genuanceerd en zijn deze antwoorden verdedigbaar correct en incorrect.

3. Hoeveel erfelijke metabole ziekten zijn er in de hielprik opgenomen?

- A. 6
- B. 18
- C. 28

Dit is een goede vraag gezien antwoorden even lang zijn en hetzelfde aspect betreffen. Bovendien zijn deze antwoorden verdedigbaar correct en incorrect. Er zit ook een duidelijk verschil tussen de antwoorden, hetgeen bij een getalsvraag belangrijk is.

4. Bij genetische hielprikscreening worden pasgeborene d.m.v. een DNA-test onderzocht op ernstige en zeldzame aandoeningen. Een voorbeeld van een DNA-test is whole genome sequencing (WGS). Wat wordt er bij WGS onderzocht?

- A. een panel van genen middels sanger sequencing
- B. alle functionele delen van het DNA
- C. het hele DNA

Dit is een goede vraag gezien opties elkaar uitsluiten en geen 'verzamelopties' omvatten.

5. Er zijn verschillende soorten innovatieve therapieën in ontwikkeling die toegepast kunnen worden bij erfelijke metabole ziekten. Welke soort therapie zou in theorie maar eenmalig hoeven worden gegeven?

- A. AAV-gentherapie
- B. mRNA therapie
- C. Small molecules therapie

Dit is een goede vraag, want de antwoorden betreffen hetzelfde domein en zijn ongeveer even lang.

6. Pyridoxine-afhankelijke epilepsie (PDE) wordt veroorzaakt door een DNA-verandering in het ALDH7A1 gen. Door deze mutatie wordt een bepaald enzym niet of onvoldoende aangemaakt. Welk enzym wordt bij PDE niet of onvoldoende aangemaakt?

- A. NAGS
- B. ASL
- C. α -AASA dehydrogenase

Dit is een goede vraag, want de antwoorden sluiten elkaar uit en betreffen hetzelfde domein. Ook de klinische relevantie is duidelijk uit de stam.

7. Patiënten met pyridoxine-afhankelijke epilepsie (PDE) kunnen worden behandeld met pyridoxine, arginine suppletie en met een dieet waarbij een specifiek aminozuur wordt vermeden. Om welk aminozuur gaat het?

- A. Leucine
- B. Lysine
- C. valine

Dit is een goede vraag, want de antwoorden betreffen hetzelfde domein, zijn even lang en sluiten elkaar uit. Ook de afdekproef (als je de opties afdekt, dan is de vraag nog steeds te beantwoorden) werkt.

Toetsvragen die bijna goed zijn

8. Patiënten met pyridoxine-afhankelijke epilepsie (PDE) **kunnen** naast epilepsie diverse andere klinische symptomen hebben. Welk klinisch symptoom **kan** bij PDE voorkomen?

- A. Hypotonie
- B. Coagulopathie
- C. Hypoglykemie

Deze vraag is bijna goed, want in de geneeskunde 'kan' alles. Beter is het om te vragen naar welk symptoom ook vaak voorkomt, of nog beter, welk symptoom het vaakst voorkomt of pathognostisch is. In dat geval werkt nl. ook afdekproef en nu is dat niet het geval.

9. Wanneer patiënten met pyridoxine-afhankelijke epilepsie (PDE) enkel met pyridoxine suppletie worden behandeld verdwijnen niet alle symptomen. Een lysine beperkt dieet biedt hier mogelijk uitkomst. Welk symptoom zou **kunnen** worden weggenomen door patiënten met PDE, naast pyridoxine suppletie, ook te behandelen met een lysine beperkt dieet?

- A. Epilepsie
- B. Groeiachterstand
- C. Ontwikkelingsachterstand

Deze vraag is bijna goed, maar moet nog iets worden aangepast, want hier speelt weer dat alles 'kan'; een casus kan hier helpen en ook door te vragen welk probleem door een dergelijk dieet het meest waarschijnlijk gunstig wordt beïnvloed.

Ongeschikte toetsvragen

1. Er zijn ruim 1.500 soorten erfelijke metabole ziekten, een aantal soorten zijn lysosomale stapelingsziekten, mitochondriële ziekten, ureumcyclus defecten en vetzuuroxidatiestoornissen. Welke van deze soorten erfelijke metabole ziekten **kunnen** cardiale klachten geven?

- A. Lysosomale stapelingsziekten, vetzuuroxidatiestoornissen en mitochondriële ziekten
- B. Lysosomale stapelingsziekten, vetzuuroxidatiestoornissen en ureumcyclus defecten
- C. Vetzuuroxidatiestoornissen, mitochondriële ziekten en ureumcyclus defecten

Deze vraag bevat dubbelingen en is derhalve geen goede vraag. Wat wel doen? Hier zou een vraag met een korter antwoord geschikter zijn. Dat kan door de dubbelingen naar de stam te halen ('naast xx en xx...'). Ook het woord 'kunnen' dient te worden aangepast (zie boven).

2. Given the following statements about recombinant VWF vs. plasma-derived VWF what/which is/are true?

- 1 Recombinant VWF is less immunogenic compared to plasma-derived VWF.
 - 2 The half-life of recombinant and plasma-derived VWF are comparable (difference in range of ≤ 30 minutes).
- a) Only statement A is true
 - b) Only statement B is true
 - c) Both statements are true
 - d) Both statements are false

Hier gaat het om 2 stellingen, waarbij de antwoordopties de 4 combi's betreft (juist/onjuist, onjuist/juist, beide juist, beide onjuist). Dat is didactisch slecht, omdat je bij een 'fout' antwoord niet weet of degene die antwoordde, niet toch een van de stellingen wel wist. Wat wel doen? Stellingvragen apart stellen als 2-keuzevragen.

3. What is the definition of prophylactic treatment in a VWD patient?

- a) Prophylaxis is defined as treatment to prevent bleeding complications during and after surgery.
- b) Prophylaxis is defined as a period of at least 6 months of treatment at least once weekly.
- c) The ISTH has not established a definition for the prophylactic treatment of VWD patients.
- d) Prophylaxis is defined as a period of at least 3 months of treatment at least twice weekly.

Vraag 3 vraagt naar een definitie; dan kan het antwoord niet zijn dat er geen definitie is (strikvraag), dus c moet weg. A is ook goed, omdat daarin geen tijdsaanduiding is en dus b en d omvat. Wat wel doen? Als het gaat om de termijn in een definitie, zou daarnaar moeten worden gevraagd (bijv. 3, 6, 9 maanden).

4. In the available studies, which outcomes were not reduced by the initiation of prophylactic treatment in VWD patients?

- a) Bleeding episodes
- b) Hospitalizations
- c) Death
- d) Heavy menstrual bleeding

Dit betreft een ontkennende (niet-)vraag. Dat is meestal klinisch niet of minder relevant. Wat wel doen? Omzetten naar een 'positieve' vraag door nog één negatieve optie te bedenken (3-keuzevraag. Eventueel kunnen 2 van de 3 positieve opties naar de stam worden verplaatst.

8. Which patients could participate in the phase 3 study of prophylactic treatment with recombinant VWF?

- a) Children
- b) VWD type 2N patients
- c) Adult patients who either received prior prophylactic treatment or prior on demand treatment.
- d) Adult patients who did not previously receive prophylactic treatment.

Het meest gedetailleerde (en langste) antwoord is meestal goed. A is te onwaarschijnlijk, omdat het veel te algemeen is en kinderen niet als patiënten benoemd zijn (die in de overige antwoorden wel staan). Als een antwoordoptie zo onwaarschijnlijk is dat deze niet wordt gekozen, kan je hem beter weglaten. Wat wel doen? Probeer antwoordopties even lang en even gedetailleerd te maken.

10. Which statement is true about the phase 3 study of prophylactic treatment with recombinant VWF?

- a) No potential adverse effects were reported during the study.
- b) Inhibitors to VWF or FVIII developed during the study.
- c) No life-threatening or fatal serious adverse events occurred.
- d) No potential adverse effects leading to discontinuation of rVWF were reported during the study

Dit is geen echte stellingvraag, al wordt in de stam wel het woord 'stelling/statement' gebruikt. Wat wel doen? Vraag hier beter naar de bijwerkingen, waar 3 van de antwoordopties over gaan (b kan dan weg, want iedereen zal meteen zien dat die er niet bij hoort en wordt dus niet gekozen).

Probeer ook hier weer te letten op gelijke lengte van de antwoordopties en detaillering. hoe korter de opties hoe beter, dus 'were reported during the study' kan weg.

Tot slot:

- Het is niet zo dat elke meerkeuzevraag 4 antwoordopties moet hebben; 2-3 kan genoeg zijn. Meer dan 5 wordt overigens afgeraden.
- Een goede vraag kan worden gemaakt zonder de antwoorden te zien (afdekproef).
- Zorg voor realistische alternatieven en maak de antwoordopties niet te lang; het is beter om de stam van voldoende informatie te voorzien, zodat de antwoordopties kort kunnen zijn.
- Maak de antwoordopties zo kort en bondig mogelijk. Men vindt een lange stam minder vervelend dan lange antwoordopties, ook omdat daarbij de afdekproef vaak niet werkt.