

Kwaliteitszorg bij toetsconstructie

De werkwijze van een multidisciplinaire centrale toetscommissie

B.H. Verhoeven, G.M. Verwijnen, A.J.J.A. Scherpbier, L.W.T. Schwirth, C.P.M. van der Vleuten

Samenvatting

Examens bepalen in sterke mate hoe studenten wat bestuderen en toetsresultaten zijn sterk bepalend voor de studievoortgang van studenten. Toetsen dienen dan ook zorgvuldig geconstrueerd te worden. Recent wordt in het visitatierapport wederom de instelling van centrale toetsbeoordelingscommissies bepleit om de kwaliteit van toetsen te bewaken. De spil van het beoordelingsproces in zulke commissies is 'peer review'. Hoewel 'peer review' zeer gebruikelijk is om de kwaliteit van onderwijsproductanten te bevorderen, is de toepassing ervan in het onderwijs uitzonderlijk. Als voorbeeld van de werkwijze van een centrale toetscommissie wordt in dit artikel de aanpak van de Maatrichse Voortgangstoets Beoordelings Commissie gedetailleerd beschreven.

Inleiding

Het doel van toetsing is een valide en betrouwbare meting van het (kennis)niveau van studenten. Fout-positieve en fout-negatieve beslissingen moeten tot een minimum beperkt blijven. Toetsresultaten hebben ingrijpende gevolgen zowel voor de studievoortgang van individuele studenten als voor het studierendement van faculteiten.¹ Bovendien bepalen examens in sterke mate hoe studenten wat bestuderen.^{2,3} Bij het maken van toetsen is dan ook uiterste zorgvuldigheid geboden. Dit vereist specifieke deskundigheid en ervaring die niet gemakkelijk te verwerven zijn.⁴⁻⁷ Weinig docenten ervaren toetsconstructie als een stimulerende bezigheid. Toch zijn er in elke onderwijsinstelling wel personen te vinden die juist dit aspect van

het onderwijs boeiend vinden en die tevens bereid zijn om zich de benodigde vaardigheden eigen te maken. Omwille van de zuiverheid van het beoordelingsproces is bovendien veel te zeggen voor een zo strikt mogelijke scheiding tussen de rollen van docent en examinator.⁸

Deze overwegingen hebben in de medische faculteit van de Universiteit Maastricht geleid tot het instellen van toetsbeoordelingscommissies, die verantwoordelijk zijn voor de beoordeling van de kwaliteiten van de toetsvragen en de samenstelling van de examens. Er is een commissie voor elk van de verschillende examenonderdelen. De commissieleden zijn docenten (stafleden) die blijk hebben gegeven van belangstelling voor toetsing. Alle voorzitters van de toetscommissies zijn lid van het Evaluatie Project Geneeskunde (EPG), dat is ingesteld voor ontwikkeling, uitvoering en onderzoek van het evaluatiesysteem.⁹ Deze projectgroep bestaat uit medici en toetspecialisten en functioneert op centraal niveau in de onderwijsorganisatie. In de loop der jaren zijn ook aan andere faculteiten van de Universiteit Maastricht toetsbeoordelingscommissies ingesteld. Er zijn ook landelijke toetsbeoordelingscommissies, zoals de Centrale Examencommissie Vaststelling Opgaven (CEVO) in Nederland, de National Board of Medical Examiners (NBME) in de Verenigde Staten en het Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) in Duitsland.¹⁰⁻¹³ Recent is in het visitatierapport wederom de instelling van centrale toetscommissies bepleit.^{14,15} Er is in de literatuur echter weinig informatie te vinden over de werkwijze van dergelijke commissies.¹⁶ Daarom wordt in dit artikel, als voorbeeld van de werkwijze van een

dergelijke commissie, gedetailleerd beschreven wat in Maastricht de aanpak is van de Voortgangstoets Beoordelings Commissie (VBC) en hoe de productie van de voortgangstoets verloopt.

De voortgangstoets

In 1976 heeft Wijnen het concept 'voortgangstoetsing' geïntroduceerd en sinds het academisch jaar 1977/1978 is de voortgangstoets een vast onderdeel van het evaluatieprogramma van de medische faculteit in Maastricht.¹⁷ De basisgedachte achter de voortgangstoets is dat het gewenste studiedrag, namelijk zelfwerkzaamheid op basis van problemen, niet verstoord mag worden door een toets. Daarom is de toets erop gericht om iedere toename van geneeskundige kennis (binnen het kader van de einddoelen van de opleiding) vast te kunnen stellen. De vragen uit de voortgangstoets moeten door een basissarts juist beantwoord kunnen worden. Voor iedere afname wordt een nieuwe toets gemaakt, deels bestaand uit nieuwe en deels uit (aangepaste) eerder gebruikte vragen. Alle studenten uit alle jaargroepen maken viermaal per jaar tegelijkertijd de voortgangstoets. Zo ontstaat er een beeld van de voortgang van de studenten op de weg naar het einddoel.^{17,22} De toetsresultaten zijn zowel voor bijsturing van individuele studenten als voor evaluatie van het curriculum zeer waardevol.^{20,23} Om redenen van praktische haalbaarheid (veel vragen en veel studenten) is gekozen voor vragenstellingen die met 'juist', 'onjuist' of '?' beantwoord kunnen worden. Iedere voortgangstoets omvat zo'n 250 vragen, verdeeld over vijftien categorieën die zijn ontleend aan de International Classification of Diseases (ICD).²¹ In de toetsblauwdruk is het aantal vragen per categorie vastgesteld.²⁴ De verdeling over deze categorieën is tot stand gekomen op basis van morbiditeitsgegevens, de omvang van de verschillende onderwerpen in algemene medische handboeken en een enquête onder de vakgroepen. Er is niet

vastgelegd hoe de vragen in elke categorie verdeeld moeten zijn over de verschillende vakgebieden.

Itembank en itemproductie

De vragen - op basissartsniveau - worden gemaakt door de vakgroepen. Deze voorzien elke vraag van een literatuurverwijzing, zodat de studenten na de toets de achtergronden van het goede antwoord kunnen opzoeken. Alle aangeleverde vragen worden, zonder nadere beoordeling, opgenomen in een centrale, deels geautomatiseerde, itembank. Elke vraag wordt in één van de vijftien categorieën ingevoerd en voorzien van een vakgroeplabel en een categorielabel. Ongeveer acht maanden voor de datum waarop een voortgangstoets moet worden afgenomen, worden tussen de 400 en 500 vragen getrokken uit de itembank. Deze trekking is random per categorie (volgens de verdeling van de toetsblauwdruk). Per categorie worden meer vragen 'getrokken' dan uiteindelijk nodig zijn, om alternatieven te hebben voor vragen die de toets der kritiek niet doortaan. Hoeveel vragen van een vakgroep bij deze trekking geselecteerd worden, is dus deels afhankelijk van het aantal vragen waarmee een vakgroep in de itembank vertegenwoordigd is. Hoe meer vragen, hoe groter de kans van die betreffende vakgroep om getrokken te worden. Een vakgroep die te weinig vragen aanlevert voor de itembank loopt dus het risico niet vertegenwoordigd te zijn in de toets. Om excessen in te genovergestelde richting (te veel vragen van een vakgroep) te voorkomen, mogen de vragen van elke vakgroep ten tijde van de trekking niet meer dan 6% procent van de itembank in beslag nemen. Hiertoe worden, voorafgaand aan de trekking, at random geselecteerde vragen 'opzij gezet' van de vakgroepen die dit percentage overstijgen. De getrokken vragen worden beoordeeld door de VBC en na goedkeuring opgenomen in de toets. In een toets gebruikte vragen die na de afname niet zijn afgekeurd, zijn pas na drie jaar weer beschikbaar voor een trek-

Tabel 1. Productiecyclus van een voortgangstoets.

Week	Omschrijving
1-2	invoer en verwerking nieuwe vragen in de itembank
3	updatng itembank en trekking van de vragen
4, 5	individuele beoordeling door VBC-leden
6-8	1e ronde van 6 VBC-besprekingen (2/wk)
9	administratieve verwerking en verzamelen en verzenden correspondentie aan vakgroepen
10-12	periode van overleg met vakgroepen
13-15	2e ronde van 6 VBC-besprekingen (2/wk)
16, 17	administratieve verwerking / voorbereiding toetsselectie door VBC-leden
17	controle mutaties en selectie van toets (alle VBC-leden)
18	administratieve verwerking / samenstellen concepttoets
19	controle concepttoets en rapportage (voorzitter en vice-voorzitter VBC)
20	administratieve verwerking / samenstellen definitieve toets
21-25	reproductie van de toets / voorbereiden van de toetsafname
26	toetsafname (woensdag 09.00u - 13.00u)
27	verwerking data en studentcommentaar / inhoudelijke voorbereiding VBC-nabespreking
28	VBC-nabespreking gevolgd door overleg met vakgroepen / berekenen definitieve resultaten
29	rapportage studenten en rapportage commissies / opstellen itemrapportage
30	itemrapportage aan de vakgroepen

king. De vragenbank is dus voortdurend in beweging. In de loop der jaren hebben alle vakgroepen tezamen ongeveer 19.000 vragen gemaakt. Sinds 1976 zijn er in de loop van het beoordelingsproces ruim 6.000 vragen afgekeurd, zodat er op dit moment circa 13.000 vragen voor toetsen beschikbaar zijn. Na iedere trekking (viemaal per jaar) krijgen de vakgroepen een overzicht van het aantal vragen waarmee zij per categorie vertegenwoordigd zijn in de itembank. Naar aanleiding van deze overzichten kan een vakgroep besluiten nieuwe vragen voor de voortgangstoets in te leveren om hun voorraad in de bank op peil te houden.

De Voortgangstoets Beoordelings Commissie

De VBC is verantwoordelijk voor het samenstellen en de kwaliteitsbewaking van de voortgangstoetsen. De commissie bestaat uit een voorzitter, een vice-voorzitter en zes leden. De voorzitter en de vice-voorzitter zijn arts en lid van het centrale toetsproject EPG. De leden vertegenwoordigen de basisvakken, klinische

vakken en gedragswetenschappelijke vakken. VBC-leden worden benoemd voor een periode van vier jaar met de mogelijkheid van een eenmalige herbenoeming. Er is expliciet gekozen voor deze relatief lange benoemingstermijn opdat maximaal gebruik gemaakt kan worden van de 'on the job' verworven vaardigheden en expertise. Voor de logistische coördinatie en administratieve ondersteuning van alle VBC-werkzaamheden is een medewerker aangesteld, die tevens verantwoordelijk is voor het beheer van de itembank en de organisatie van de toetsafname.

Werkwijze van de VBC

Er wordt gewerkt volgens een strak productieschema waarin alle activiteiten en bijbehorende tijdstippen precies zijn vastgelegd. Alle documenten die bij de vraagproductie en kwaliteitsbewaking betrokken zijn, worden aan het begin van elk academisch jaar geïnformeerd over de productiedata. De totale productiecyclus van één voortgangstoets omvat een periode van dertig weken (tabel 1). De VBC

heeft tweemaal per week werkbijeenkomsten. Deze zijn steeds op een vast tijdstip in de week gepland en duren circa drie à vier uur. Omdat er vier toetsen per jaar worden geproduceerd, werkt de VBC gelijktijdig aan verschillende toetsen. Dit vereist een goede planning om de productieschema's goed op elkaar af te stemmen en piekbelastingen door het jaar heen zo veel mogelijk te voorkomen.

Beoordeling vóór toetsafname

De 400 à 500 vragen die per ronde uit de itembank getrokken worden, gaan naar de VBC. De leden van de VBC zijn elk verantwoordelijk voor één of meer (maximaal vier) categorieën van de blauwdruk. De beoordeling van de geselecteerde toetsvragen gebeurt in eerste instantie door ieder VBC-lid afzonderlijk. Ieder controleert de vragen van de eigen categorie(en). De bevindingen worden vervolgens in de volgende commissie besproken en gecontroleerd om elkaar van feedback te kunnen voorzien. Hier is een eerste ronde van zes besprekingen voor gereserveerd. De besprekingen vinden plaats in een ruimte waar een groot aantal boeken ter beschikking is om de inhoudelijk gelidigheid van de vragen te kunnen controleren. Bij de beoordeling van de vragen wordt niet alleen nagegaan of de formulering van de vraag voldoende eenduidig is, maar ook of de vraag inhoudelijk klopt en eenduidig gedocumenteerd is. Als een vraag in een van beide opzichten niet deugt, worden concrete redactiewijzigingen voorgesteld. Hierbij wordt altijd de contactpersoon van de vakgroep die de vraag inleverde via schriftelijk en/of mondeling overleg betrokken. Zonder overleg met de desbetreffende vakgroep kunnen vragen niet aanangepast of verwijderd worden, omdat de inhoudelijke kwaliteit van de vragen onder de verantwoordelijkheid van de vakgroepen valt. Het resultaat van het overleg met de contactpersoon wordt wederom plenair in de VBC besproken in een tweede ronde van zes besprekingen. Naar aanleiding van deze bespreking kan

een vraag goedgekeurd of geëlimineerd worden, of er kan opnieuw overleg nodig zijn. Doordat voor veel vragen meer dan één overlegronde nodig is, en er tegelijkertijd aan verschillende toetsen gewerkt wordt, heeft de VBC continu ongeveer duizend vragen in behandeling.

Ten slotte is er nog een laatste bijeenkomst waarin de betreffende voortgangstoets in concept wordt samengesteld. Dit doet elk VBC-lid voor de eigen categorie(en). In deze fase wordt gecontroleerd of alle aanpassingen correct verwerkt zijn en wordt een keuze gemaakt uit de beschikbare verzameling goedgekeurde vragen. De medewerking van de vakgroepen tijdens de voorbereidingsfase van elke toets heeft dus (naast het aantal vragen in de itembank) invloed op het uiteindelijke aantal vragen dat per vakgroep in de toets wordt opgenomen. Een inadequate reactie op het VBC-commentaar beperkt het aantal vragen dat uiteindelijk beschikbaar is voor opname in een voortgangstoets. Elk VBC-lid selecteert een klein surplus aan vragen voor de concepttoets, opdat er in de slotfase van de toetsaansameling nog vragen verwijderd kunnen worden om een evenwichtige verdeling over de antwoordsleutels en de drie vakkenclusters (basisvakken, klinische vakken en gedragswetenschappelijke vakken) te bewerkstelligen. Na de administratieve verwerking wordt de concepttoets in zijn totaliteit nagekeken door de voorzitter en vice-voorzitter. Oorspronkelijk was deze fase bedoeld voor een laatste controle op tikfouten en overlap van vragen. De ervaring leerde echter dat er in dit stadium nog steeds essentiële problemen ontdekt werden. Zodoende vindt er tegenwoordig in deze fase ook nog een controle plaats op verkeerde formuleringen, literatuurverwijzingen en inhoudelijke fouten. De bevindingen van deze controle worden schriftelijk en op naam gerapporteerd aan de leden van de VBC en plenair nabesproken, als een soort interview. Verder worden de sleutelverdeling (verhouding juist/onjuist) en de clusterverdeling in de

concepttoets berekend. Door vragen te verwijderen wordt gestreefd naar een evenredige sleutelverdeling (circa 50% juist / 50% onjuist) en de gewenste clusterverdeling (40% basisk vakken / 40% klinische vakken / 20% gedragswetenschappelijke vakken). Als laatste stap voordat de toets gedrukt wordt, neemt de voorzitter de gehele toets nog eens door om te controleren of alle aanpassingen goed zijn verwerkt.

Van alle vragen die uit de itembank worden getrokken (deels nieuw en nog nooit beoordeeld, deels eerder gebruikt) blijken de meeste (ruim 80%) niet meteen geschikt om in een toets op te nemen. In vrijwel alle gevallen (97%) wordt bij deze vragen een vormprobleem gesignaleerd. Bij meer dan de helft (56%) is er echter ook sprake van een inhoudelijk probleem.^{25,26}

Beoordeling na toetsafname

Na afname van een voortgangstoets worden de vragen opnieuw beoordeeld. Ondanks de nauwgezette beoordeling vooraf verschijnen toch ondeugdelijke vragen in de toets. Zowel de vakgroepen als de VBC zien problemen over het hoofd. De VBC kan op dergelijke misers geattendeerd worden door de studenten, die expliciet als partner betrokken worden in het kwaliteitsbewakingsproces.^{27,29} Hiertoe kunnen studenten tot één week na de toetsafname commentaar indienen. Vanwege problemen met de leesbaarheid van geschreven commentaar wordt alleen getypt commentaar geaccepteerd. Als de student literatuurverwijzingen gebruikt, moeten hiervan kopieën bijgevoegd zijn. Er wordt gemiddeld door 4% van de studenten commentaar geleverd op ongeveer 20% van het aantal vragen van een toets.

Om probleemvragen op te sporen wordt tevens gebruik gemaakt van de statistische parameters van de toetsresultaten per item, de itemanalyse.³⁰ Zowel de studentcommentaren als de resultaten van de itemanalyse worden toegevoegd aan een tekstbestand waarin ook

de tekst van de vragen is opgenomen. De voorzitter en vice-voorzitter bekijken elke vraag aan de hand van de studentcommentaren en de itemstatistieken. De studentcommentaren worden inhoudelijk beoordeeld door de meegestuurde literatuur te controleren. Indien het studentcommentaar steekhoudende argumenten bevat, wordt de betreffende vraag geselecteerd voor een plenaire bespreking in de VBC, de nabespreking. Ook alle vragen met een curieus antwoordpatroon (geen groei over de jaren, een opvallende piek of dal in een bepaalde jaargroep, door meer dan 30% van de zesdejaars fout beantwoord of door meer dan 50% van de zesdejaars niet beantwoord) worden geselecteerd voor deze nabespreking. Anderhalve week na de toetsafname komt de voltallige VBC hiervoor bijeen. Alle geselecteerde vragen worden, voorzien van de voorgeschiedenis en een behandelingsadvies van de voorzitter en vice-voorzitter, besproken. Tijdens de vergadering wordt besproken of vragen teruggetrokken of antwoord sleutels gewijzigd moeten worden. Indien de VBC eliminatie of sleutelwijziging overweegt, wordt de vakgroep waarvan de vraag afkomstig is voor nadere inhoudelijke beoordeling geconsulteerd. In overleg met de vakgroep wordt vastgesteld of de vraag onveranderd gehandhaafd blijft, er een sleutelwijziging doorgevoerd moet worden of de vraag moet komen te vervallen. Er wordt geen enkele vraag verwijderd of antwoord sleutel gewijzigd zonder instemming van de vakgroep die de vraag gemaakt heeft. Nadat dit overleg met de vakgroepen heeft plaatsgevonden, worden de definitieve toetsresultaten berekend. De vervallen vragen worden bij het berekenen van de studentscores buiten beschouwing gelaten. Gemiddeld valt 5,5% van de vragen, maar dit percentage kan per toets behoorlijk variëren met uitersten van 0,4%-11,5%. Het aantal sleutelwijzigingen is gemiddeld 0,7% en varieert tussen de 0% en 2%. Na verwerking van de sleutelwijzigingen en het verwijderen van geëlimineerde vragen blijkt dat de toetsen toch altijd nog vra-

gen bevatten die door relatief weinig studenten ingevuld of door relatief veel studenten fout beantwoord worden. Op het einde van het zesde jaar (het 24ste meetmoment) wordt gemiddeld zo'n 15% (circa 34 vragen) van de vragen door minder dan 50% van de zesdejaars studenten beantwoord. Er zijn wat dit betreft echter aanzienlijke verschillen tussen de voortgangstoetsen. Het aantal vragen dat door meer dan 30% van de zesdejaars fout beantwoord wordt, ligt daarentegen vrij stabiel rond de 16% (circa 36 vragen).

Rapportages

De resultaten worden gerapporteerd aan de Examencommissie, die de uitslagen formeel vaststelt en bekendmaakt aan de studenten. De studenten krijgen zowel hun eigen individuele resultaat als de gemiddelde resultaten van de jaargroep waartoe ze behoren. Er wordt een overzicht gegeven van de score per categorie en per vakgebied. De scores worden weergegeven als percentages van het maximaal te behalen resultaat. Zowel de goed-, fout-, vraagtekenscores als de goed-min-foutscore worden aangegeven op het formulier. Naast de individuele score staat middels '+' en '-' aangegeven of het resultaat hoog respectievelijk laag is in vergelijking met de jaargroep. Onder dit overzicht staan de totaalscore (goed-min-fout) en de kwantificatie (voldoende/onvoldoende). Tevens is op dit formulier aangegeven welke vragen zijn vervallen en welke antwoordsleutels gewijzigd zijn (figuur 1).

De resultaten worden ook routinematig gerapporteerd aan de Opleidingscommissie, EPG en alle vakgroepen. Elke instantie krijgt een overzicht van de gemiddelde resultaten per jaargroep op de toets als geheel en uitsplitsend naar de categorieën, vakgroepen en clusters van vakgroepen waar de betreffende vragen van afkomstig waren. Bij deze rapportage worden de zogenaamde groeicurveën plus de resultaten van het betreffende academisch jaar gevoegd (figuur 2).²¹

Daarnaast ontvangt elke vakgroep een specifieke itemrapportage van al hun vragen die in de betreffende toets waren opgenomen. Naast de definitieve tekst wordt per vraag het antwoordprofiel, eventueel studentcommentaar en geschiedenis (eerdere versies, resultaten van voorgaande afnames, eventuele sleutelwijziging, het vervallen na afname; enzovoorts) gerapporteerd. Feedback wordt gegeven in de vorm van de itemanalyse, eventueel studentcommentaar en opmerkingen van de VBC. Deze opmerkingen bestaan uit een korte interpretatie van de itemanalyse en mogelijke oorzaken van gesignaleerde problemen. De studenten zijn in de gelegenheid deze zogenaamde itemrapportage in te zien aan de onderwijsbalie (figuur 3).

Beroepsprocedure

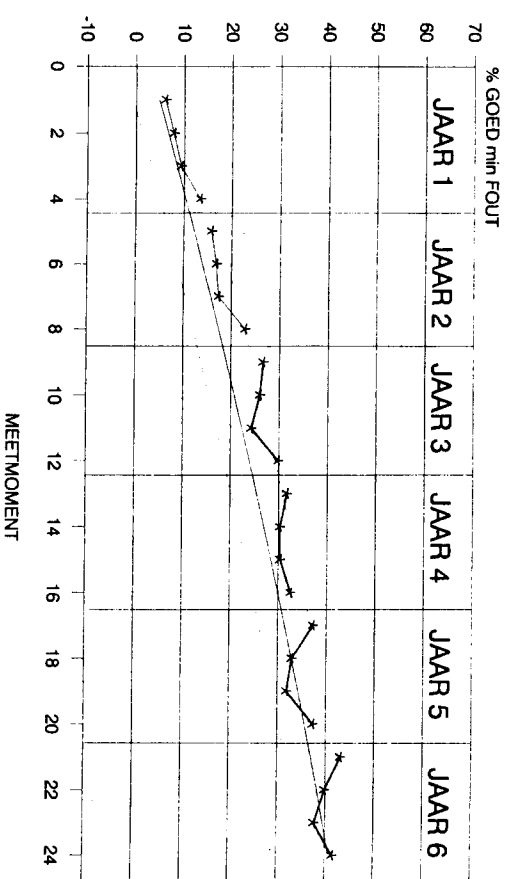
Nadat de uitslag is vastgesteld kan een student die het oneens is met de wijze waarop die uitslag totstandgekomen is, beroep aantekenen bij het College van Beroep voor de Examens (CBE).³¹ Deze mogelijkheid is wettelijk voorgeschreven.³² De voorzitter van de VBC verzorgt dan een schriftelijk commentaar met advies over het beroep aan de voorzitter van de Examencommissie, zo nodig na consultatie van betrokken inhoudsdeskundigen. Op basis hiervan wordt een procedureel voorgeschreven overleg gevoerd met de student om te bepalen of een zogenaamde minnelijke schikking te bereiken is. Indien partijen geen overeenstemming bereiken, dan wordt de zaak in een openbare zitting van het CBE behandeld, al dan niet in aanwezigheid van opgeroepen getuigedeskundigen. Sinds 1990 wordt regelmatig een beroepszaak aangespannen. Voorheen was dit slechts eenmaal eerder voorgekomen, vermoedelijk omdat pas sinds maart 1990 de beroepsprocedure mogelijk werd vermeld op het officiële uitslagformulier. Voor die tijd werd de mogelijkheid uitsluitend beschreven in de studiegids. In totaal is er in zeven jaar (1990-1996) te-
gen 121 afzonderlijke vragen beroep

Categorie (omschrijving)	aantal vragen	RESULTATEN PER CATEGORIE (procentueel)		studiegroep (n=215)	
		Goed	Fout	Goed	Fout
1 Ademhalingsmedicatie	28	50++	7--	43	43++
2 Hereditair en perinatale	14	36	14	50	21
3 Spijt- en skelet	18	44	22+	33--	22
4 Geneeskundige gez-zorg	17	53-	29+-	18	24--
5 Voortplantingsstelsel	16	13-	31-	56-	-19--
6 Hart- en vaatstelsel	28	71++	14	14--	57++
7 Hormoonstelsel en metabolisme	16	19-	13	69-	6
8 Huid en bindweefsel	14	50+	0-	50+	31
9 Misdiagnose	14	64	14	21	50
10 Spijvertrappingsstelsel	23	11-	17	72+	-6-
11 Nieren en urinewegen	18	33	14	52	73++
12 Zenuwen en zintuigen	21	9--	18	73++	-9--
13 Kest	11	56+	22	22	33
14 Wetenschap en methode	9	50++	0--	0--	0--
15 Pess fact ziekevgez	6	50	16	45	39
Totaal	249	39	16	45	39

Vakgroep (omschrijving)	aantal vragen	RESULTATEN PER VAKGROEP (procentueel)		studiegroep (n=215)	
		Goed	Fout	Goed	Fout
1 Anatomie/Embryologie	10	30-	7-	50+	10-
2 Biochemie	15	33	20-	60	27-
3 Biofysica	0	0	0	0	0
4 Farmacologie	10	50+	20	30-	30
5 Fysiologie	20	55-	5--	40+	50++
6 Genetica/Anthropologie	10	0-	0-	70++	-30--
7 Immunologie	9	22-	11+	67+	11--
8 Medische microbiologie	18	6--	11+	83+	-6-
9 Pathologie	9	33	11	56	22
Basisvakken	101	30-	13-	57+	17
10 Algemene heekunde	10	60++	10	30-	50++
11 Cardiologie	12	67++	8-	25	58++
12 Dermatologie	9	14+	0-	56	44++
13 Gynaecologie/Obsistrie	10	20-	20-	30-	-30--
14 Histologie	10	40	20	40	20-
15 Inwendige geneeskunde	14	14-	7-	79-	7
16 Kinder geneeskunde	9	33	22	44-	11
17 Kede-,Neus-,Oorheik	5	80++	0--	20-	80++
18 Neurologie	5	20	20	60	0
19 Oogheelkunde	4	25	25+	50	0
20 Orthopedie	3	0	0	33+	67-
21 Pulmonologie	3	33	33	67-	-33--
22 Radiodiagnostiek	3	67++	33	0-	33+
23 Revalidatiegeneeskunde	1	1	0	100+	0
24 Urologie	2	0-	0	100+	0
Klinische vakken	106	38	16	46	22+
25 Ehfkd/Emcn/Cg/zorg	2	50	50	0-	0
26 Epidemiologie	4	50	0-	50+	50+
27 Geriatrie	7	71	14	14	67
28 Wetensc ethiek	5	80++	20	0-	60+
29 Medische psychologie	10	70++	30++	0-	40
30 Medische sociologie	5	40-	60++	0-	-20--
31 Psychiatrie	9	56	21	33	44
Cedraswavel vakken	42	62	21++	45	38
Totaal	249	39	16	45	39

De norm voor deze tests is: 15,85
 Uw procentuele GOED-FOUT score is: 22,49 (Absoluut: 56)
 Niet opgenomen vragen: 38,68,69,133,152,155 en 242
 Dit resultaat is VOLDOENDE
 Studeerwijzigingen: 12,11,6,21,5 en 243

Figuur 1. Voorbeeld van een uitslagformulier zoals de studenten het ontvangen. Op de bovenste helft worden de individuele score en jaargroepscore per categorie weergegeven en direct eronder per vakgroep.



Figuur 2. Groeicurve gebaseerd op voortgangstoetsresultaten van september 1985 tot en met mei 1995. De curve representeert de 'line of best fit' door de gemiddelde jaargroepscores van 10 cohorten studenten (10 maal 24 datapunten). De stippe lijnen representeren de 95% betrouwbaarheidsintervallen. Met sterren zijn de jaargroepgemiddelden van de vier toetsen van het academisch jaar 1995-1996 weergegeven.

aangekend, bij 19 van de in totaal 28 toetsen. Relatief gezien is maar een klein aantal vragen aangevochten. Het aantal van 121 vragen is nog geen 2% van het totale aantal gestelde vragen in deze zeven jaar. Het merendeel van de bezwaren werd in de zogenaamde schikking-fase van de beroepsprocedure aangehandeld. Bij 27 van de 121 aangevochten vragen konden student en Examencommissie het niet eens worden en was een uitspraak van het College van Beroep voor de Examen nodig. Beroepszaken worden aangekaart door studenten die daar een specifiek belang bij hebben. In totaal betrof het in de afgelopen zeven jaar 59 verschillende studenten, die vrijwel allemaal gebaat waren met een kleine verbetering van hun resultaat om alsnog een betere kwalificatie te krijgen.

Beschouwing

Toetsbeoordelingscommissies zijn bedoeld om onafhankelijk van de auteurs te controleren of toetsvragen valide indicatoren voor kennis zijn. Door storingsbronnen op te sporen wordt geprobeerd de kans op fout-positieve en fout-negatieve beslissingen te minimaliseren, zodat de uiteindelijke toets zo goed mogelijk meet wat beoogd wordt te meten. Het blijkt zinvol om inhoudelijke en vormtechnische problemen te signaleren in een groep waarin verschillende disciplines vertegenwoordigd zijn. Het in dit artikel beschreven kwaliteitszorgproces kan beschouwd worden als een vorm van 'peer review'. De werkwijze van de VBC ligt niet vast en wordt voortdurend aan de geldende omstandigheden, eisen en opvattingen aangepast. De beschreven manier van werken is in de loop van de afgelopen twintig jaar door voortdurende aanpassingen ontstaan en wordt continu bij-

VOORGAANGSTOETS H&G-4L December 1994		EVALUATIE-PROJECT GENESKINDJE					
FYS 0134 06-00195							
De doorstroming van het coronarystroom is tijdens de diastolische fase van de hartcyclus anders dan tijdens de systolische fase.							
105-	In de diastolische fase is de doorstroming het grootst.						
SI:	Just						
ITEMANALYSE:							
	STUDIEER:	1	2	3	4	5	6
	%-GOED:	11	38	58	64	71	79
	%-FOUT:	8	21	28	20	20	15
	%-VRGT:	82	21	14	16	9	6
	RIJ Gmf:	0,089	0,001	0,118	0,104	0,093	0,216
	DI Gmf:	-0,037	-0,026	0,048	0,133	0,026	0,160
STUDENTCOMMENTAAR:							
#93152							
De doorstroming van het coronarystroom is tijdens de hartcyclus aan sterke veranderingen onderhevig. De linker coronairarterie wordt tijdens de systole door de hoge intraventriculaire druk dichtgedrukt (vnl. de binneste capillaren en vaten). De sterkste omringing vindt hier dus wél tijdens de diastole plaats. De rechter coronairarterie heeft niet zo'n last van de intraventriculaire druk, omdat deze rechts veel lager is dan links. Dit betekent dat de doorstromingsnelheid de normale volgg. dus deze doorstroming is het grootste tijdens de systole. Aangezien bij 50% van de mensen de rechter coronairarterie dominant is (20% = linksdominant, 30% = gelijke verdeling) - dit betekent dat deze arterie de rechterkant en een gedeelte van de linker kant verzorgt - zal dus bij 50% van de mensen 'het coronarystroom' voor het grootste gedeelte uit de rechter coronairarterie bestaan, en dus het sterkst doorbloed worden.							
Lic. Bernards en Bouman, Fysiologie van de mens, 1988, blz 361.							
#XXXX8							
Volgens mij is dit een discussabele vraag, omdat je de doorstroming van het coronarystroom niet even kan kunt scheren. De linker coronair arterie heeft inderdaad de beste doorstroming aan het begin van de diastole en vrijwel geen doorstroming tijdens de systole. ECHTER de rechter coronair arterie volgt de druk van de aorta, de doorstroming is daar dus maximaal op het einde van de systole.							
(Bernards & Bouman 1988, pag. 361)							
OPMERKINGEN VOOR GAANGSTOETS BEOORDELINGS COMMISSIE (VBC).							
De vraag kwam in overleg met een contactpersoon uit de vakgroep na toetsaanname te VERVALLEN.							
Motivering: Het studentcommentaar is correct en terecht.							
Wilt men jullie de vraag nog reviseren of moet de vraag nu definitief uit het voortgangstoetsbestand verwijderd worden?							
Voorschrijft:							
Oorspronkelijke versie (gebruikt in VT-DEC87; vergelijkbare respons, toen minder FOUT; desijds meer four-postieve?)							
De doorstroming van het coronarystroom is tijdens de diastolische fase van de hartcyclus groter dan tijdens de systolische fase [juist]							

Figuur 3. Een voorbeeldvraag uit een itemrapportage met studentcommentaar, itemanalyse en opmerkingen van de (vice-)voorzitter.

gesteld. Zo heeft het verschijnen van het Raamplan 1994 artsopleiding recent weer geleid tot voorstellen om de toetsvragen nauwer aan te laten sluiten bij dit Raamplan.³³⁻³⁶ Ook wordt momenteel nagegaan of de werkwijze door onder andere (her)automatisering efficiënter gemaakt kan worden.

Onze ervaring leert dat de beschreven toetsbeoordeling veel problemen aan het licht brengt, die tijdig opgelost kunnen worden. Van de vragen die getroffen zijn uit de item-

afgelopen twintig jaar is dit percentage constant gebleven. Hetzelfde geldt voor andere psychometrische parameters. Over de jaren is de betrouwbaarheid, de gemiddelde jaargroepscore en het percentage vragen dat niet of fout wordt ingevuld stabiel.^{21, 37}

Het is niet aangetoond dat de hier beschreven werkwijze van de centrale toetscommissie tot betere toetsvragen leidt. De werkwijze van de VBC betekent echter wel dat er een helder, grondig en openbaar beoordelingsproces is, waardoor alle betrokkenen weten dat er serieus met de toets omgegaan wordt. Vanwege de consequenties van toetsen voor de voortgang van studenten mogen studenten die zorg voor de kwaliteit verwachten. Hoewel 'peer review' zeer gebruikelijk is om de kwaliteit van onderzoeksproducten te bevorderen, is de toepassing ervan in het onderwijs uitzonderlijk. Onzes inziens is de tijd rijp om 'peer review' ook in het onderwijs intensiever toe te passen.

Literatuur

- Cohen-Scholimus J, Vleuten CPM van der, Bender W. Een betere cesuur bij tentamens. Onderzoek van Onderwijs 1996;25(3):54-5.
- Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. Med Educ 1983;17:165-71.
- Frederiksen N. The real test bias: Influences of testing on teaching and learning. American Psychologist 1984;39(3):193-202.
- Ebel RL, Frisbie DA. Essentials of educational measurement. (5th ed.) New Jersey: Englewood Cliffs, 1991.
- Haladyna TM, Downing SM. A taxonomy of multiple-choice item-writing rules. Applied Measurement in Education 1989;2(1):37-50.
- Cox KR. How to construct a fair multiple choice question paper. In: Cox KR, Ewan CE, redactie. The medical teacher. (2nd ed.) London: Churchill Livingstone, 1988:157-60.
- Eekhof JAH, Pollemans MC, Verwijnen GM. Toetsbeoordeling. In: Metz JCM, Schepfler AJJA, Vleuten CPM van der, redactie. Medisch onderwijs in de praktijk. Assen: Van Gorcum, 1995:232-7.
- Wijnen WHFW. Bouwstenen voor een nieuwe medische faculteit. Bulletin Medisch Onderwijs 1997;16(1):55-64.
- Verwijnen GM, Imbos T. Ontwikkelingsplan van de Deelprojectgroep Summatieve Evaluatie. Maastricht: Rijksuniversiteit Limburg, Faculteit der Geneeskunde, 1979.
- Stokes JF. Examining in the United States: The National Board of Medical Examiners. Br J Med Educ 1967;1:320-9.
- Hubbard JP. Measuring medical education. (2nd ed.) Philadelphia: Lea & Febiger, 1978.
- Kraemer HJ, Dupré HJ, Boelcke G, Micaelis J, Voigtmann K. Institut für medizinische und pharmazeutische Prüfungsfragen (IMP). Aufgaben, Entwicklung, Analysen. Mainz: Verlag Druckhaus Schmidt & Bödige, 1976.
- Kreft HPI. Examen: VBO/MAVO/HAVO/VWO 1995. Amhem: CITO, 1995.
- VSNU-Visitatiecommissie Geneeskunde en Gezondheidswetenschappen. Onderwijsvisitatie Geneeskunde en Gezondheidswetenschappen. Utrecht: Vereniging van Samenwerkende Nederlandse Universiteiten, 1992.
- VSNU-Visitatiecommissie Geneeskunde en Gezondheidswetenschappen. Onderwijsvisitatie Geneeskunde en Gezondheidswetenschappen. Utrecht: Vereniging van Samenwerkende Nederlandse Universiteiten, 1997.
- Downing SM, Haladyna TM. Test item development: validity evidence from quality assurance procedures. Applied Measurement in Education 1997;10(1):61-82.
- Wijnen W. Einddoel-toetsen: waarom en hoe? Onderzoek van Onderwijs 1977;6(3):16-9.
- Verwijnen GM, Imbos T. Evaluatie aan de medische faculteit Maastricht. Metamedica 1978;57(1):21-32.
- Wijnen WHFW, Vleuten CPM van der. Toetsing: hordelooop of voortgangskontrolle? Universiteit en Hogeschool 1985;31(6):270-9.
- Verwijnen GM, Pollemans MC, Wijnen WHFW. Voortgangstoetsing. In: Metz JCM, Schepfler AJJA, Vleuten CPM van der, redactie. Medisch onderwijs in de praktijk. Assen: Van Gorcum, 1995:225-31.
- Vleuten CPM van der, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. Med Teach 1996;18(2):103-9.
- Vleuten CPM van der, Schepfler AJJA, Wijnen WHFW, Snellen HAM. Flexibility in learning: a case report on problem-based learning. International Higher Education 1996;2:17-24.
- Verwijnen GM. Betekenis van studieresultaten bij studiebegeleiding. Een handvol ervaringen. In: Grave WS de, Nuy HJP, redactie. Leren studeren in het hoger onderwijs: perspectieven voor integratie. Almere: Versluis Uitgeverij bv., 1987:95-110.

24. Verwijnen M, Imbos T, Snelten H, Stalenhoef B, Pollemaans M, Luyk S van, et al. The evaluation system at the medical school of Maastricht. Assessment and Evaluation in Higher Education 1982;7(3):223-44.
25. Hessen PAW van, Verwijnen GM. Toetsen getoetst. Het beoordelen van toetsvragen in Maastricht. Bulletin Medisch Onderwijs 1989;8(3):100-5.
26. Verwijnen GM. Jaarverslag 1994 Voortgangstoets Beoordelings Commissie. Maastricht: Universiteit Maastricht, Faculteit der Geneeskunde, 1995.
27. Bandranajayake RC, Cox KR. Writing multiple choice questions. In: Cox KR, Ewan CE, redacte. The medical teacher. (2nd ed.) London: Churchill Livingstone, 1988;152-6.
28. Verwijnen GM. De student als kwaliteitsbewaker. De rol van de student bij de kwaliteitsbewaking van de Maastrichtse voortgangstoets. Bulletin Medisch Onderwijs 1994;13(2):87-95.
29. Prince CIAH, Visser K. The student as quality controller. In: Scherphier AJA, Vleuten GPM van der, Reihans JJ, Steeg AFW van der, redacte. Advances in medical education. Dordrecht/ Boston/ London: Kluwer Academic Publishers, 1997;15-8.
30. Bender W. Item-analyse. Bulletin Medisch Onderwijs 1994;13(1):37-43.
31. Examencommissie Faculteit der Geneeskunde. Artikel 33 van de Regelingen voor onderwijs en examens. In: Studiegids 1997/1998. Faculteit der Geneeskunde. Maastricht: Bureau Onderwijs Faculteit der Geneeskunde, 1997.
32. Wet Modernisering Universitaire Bestuursorganisatie. Staatsblad van het Koninkrijk der Nederlanden 1997;117:1-31.
33. Metz JCM, Pels Rijken-van Epp Taalman Kip EH, Van den Brand Valkenburg BWM. Raamplan 1994 artsopleiding, eindtermen van de artsopleiding. Nijmegen: Universitair Publikatiebureau, 1994.
34. Opleidingscommissie. Visuifterrapport VBC. Maastricht: Universiteit Maastricht, Faculteit der Geneeskunde, 1995.
35. Opleidingscommissie. De toets der kritiek: Een rapport van de Commissie Toetsstelsysteem Kennis/Inzichtdomein. Maastricht: Universiteit Maastricht, Faculteit der Geneeskunde, 1995.
36. Commissie Blauwdruk en Organisatie Voortgangstoets. Eindrapportage. Maastricht: Universiteit Maastricht, Faculteit der Geneeskunde, 1996.
37. Swanson DB. Review of the assessment system used by the University of Limburg Medical School. Maastricht, The Netherlands: Faculty of Medicine, 1988.

DE AUTEURS

B.H. Verhoeven, arts-onderzoeker en destijds ad interim voorzitter van de Voortgangstoets Beoordelings Commissie, Skillslab.

G.M. Verwijnen, huisarts, ex-voorzitter van de Voortgangstoets Beoordelings Commissie, Skillslab.

Dr. A.J.J.A. Scherphier, arts, hoofd van het Skillslab.

L.W.T. Schuurwirth, arts, ad interim voorzitter van de Voortgangstoets Beoordelings Commissie, Vakgroep Onderwijsontwikkeling en Onderwijsresearch.

Prof. dr. C.P.M. van der Vleuten, hoogerleraar onderwijskunde, projectleider Evaluatie Project Geneeskunde, vakgroepvoorzitter Onderwijsontwikkeling en Onderwijsresearch.

Alle auteurs zijn verbonden aan de Faculteit der Geneeskunde van de Universiteit Maastricht.

Correspondentieadres:

B.H. Verhoeven, Skillslab, Universiteit Maastricht, Postbus 616, 6200 MD Maastricht.

Evaluatie van onderwijs aan bedrijfsartsen-in-opleiding Over de richtlijn sociaal-medische begeleiding bij rugklachten

P.B.A. Smits, W.E. van der Weide, J.H.A.M. Verbeek

Samenvatting

Bij een opleidingsgroep van de Corvubedrijfsartsenopleiding Amsterdam werd het onderwijs in de richtlijn over de sociaal-medische begeleiding van werknemers met rugklachten onderzocht. Na het onderwijs bleek de score op een kennisstoets over de richtlijn significant verbeterd, evenals de score op de performance-indicatoren welke een maatvormen voor de toepassing van de richtlijn in de praktijk. De toepassing van performance-indicatoren lijkt bruikbaar als middel om feedback te geven aan cursisten en als evaluatiemaat voor het onderwijs. Deze indicatoren geven een indruk van wat bedrijfsartsen kunnen (performance), Geconcludeerd wordt dat het onderwijs waarschijnlijk heeft bijgedragen aan de verbetering van de kennis van de bedrijfsartsen-in-opleiding en aan een betere toepassing van de richtlijn in hun praktijk.

Inleiding

De Corvu-bedrijfsartsenopleiding Amsterdam - een samenwerkingsverband van het Coronel Instituut van het Academisch Medisch Centrum / Universiteit van Amsterdam en de Afdeling Sociale Geneeskunde van de Vrije Universiteit Amsterdam - verzorgt de postdoctorale beroepsopleiding tot bedrijfsarts. Dit is een vierjarige vervolgopleiding na de basissartopleiding, die bestaat uit vier jaar praktijkopleiding in een Arbodienst, 96 Corvu opleidingsdagen, twaalf weken stage en het schrijven van een literatuurstudie en onderzoeksverslag. Opleidingsdagen, stage en onderzoek vragen van

de bedrijfsartsen-in-opleiding gemiddeld een inzet van ruim een dag per week.

De Corvu is een goed functionerend opleidingsinstituut met zeventien jaar ervaring in onderwijs aan bedrijfsartsen. Over de taken van de bedrijfsarts en over specifieke aspecten van consultvoering door bedrijfsartsen is eerder gepubliceerd.¹⁻² Op het Coronel Instituut wordt ook onderzoek gedaan naar de effectiviteit van sociaal-medische begeleiding. Zo is onderzoek verricht naar de taken van de bedrijfsarts op het gebied van rugklachten.³⁻⁴ Beide onderzoekslijnen zijn gecombineerd in een onderzoeksproject betreffende de effectiviteit van onderwijs in sociaal-medische begeleiding aan bedrijfsartsen-in-opleiding. Het in dit artikel beschreven onderzoek is een evaluatie van het onderwijs in de sociaal-medische begeleiding van werknemers met rugklachten. Het onderwijs over de richtlijn rugklachten bestaat anderhalve opleidingsdag binnen een zesdaagse module over sociaal-medische begeleiding. Het onderwijs bestaat uit een mondelinge toelichting op de richtlijn, die eerder op schrift is verstrekt, en die de bedrijfsartsen-in-opleiding van tevoren hebben kunnen bestuderen. Daarnaast wordt aandacht besteed aan de standaarden betreffende rugklachten van het Nederlands Huisartsen Genootschap en aan de mogelijkheden tot verwijzing naar bijvoorbeeld een naschool. Tenslotte wordt feedback gegeven naar aanleiding van de casus die de bedrijfsartsen-in-opleiding van tevoren hebben ingeleverd.

Dit onderzoek is bedoeld als een eerste stap in het verkrijgen van inzicht in de effectiviteit van onderwijs in sociaal-medische begeleiding. Omdat bedrijfsartsen-in-opleiding voor