

# Programma-evaluatie in relatie tot toetsprestaties: een case-study

H. van Mameren, H.A.P. Wolfhagen, D. Riksen

## Inleiding

Het onderwijs in jaar 1 tot en met 4 aan de Faculteit der Geneeskunde van de Rijksuniversiteit Limburg is opgedeeld in blokken van zes weken. Gedurende deze periode werken studenten in onderwijsgroepen aan de hand van casusmateriaal dat is opgezet rondom een bepaald thema, bijvoorbeeld Bewegen en belasten (blok 2.2), Pijn (blok 3.5) en Bloedverlies (blok 4.5). Elke onderwijsgroep wordt begeleid door een tutor. Dit is een docent uit de vaste wetenschappelijke staf, die hiervoor een speciale 'tutortraining' heeft gevolgd. Uitgaande van de doelstellingen van het blok stelt de Studierichting Commissie (SRC) vast welke vakgroepen als kerndisciplines in een blok participeren. De Onderwijscoördinatie Commissie (OCC) benoemt uit deze kerndisciplines een blokcoördinator en een planningsgroep. Deze zijn verantwoordelijk voor de uiteindelijke vormgeving van het blok, zoals de uitwerking van de casuïstiek en het vaststellen van de onderwerpen voor colleges en practica. Daarnaast verzorgt het skillslab aanvullende vaardigheidstrainingen. Elk blok wordt afgesloten met een bloktoets. De planningsgroep is verantwoordelijk voor de samenstelling van deze bloktoets. Een bloktoets heeft betrekking op de leerstof van het blok en bestaat uit ongeveer 150 gesloten vragen van het juist/?/onjuist type. De eindscore op de bloktoets wordt uitgedrukt in de Goed-minus-Fout score.

Om inzicht te krijgen in de kwaliteit van het onderwijs, wordt de studenten gevraagd een oordeel te geven over het gevolgde onderwijsprogramma. Hiertoe vullen zij aan het einde van elk blok een vragenlijst programma-evaluatie in (doorgaans na afloop van de toets, soms tijdens de laatste onderwijsgroepsbijeen-

komst). Het project Programma-Evaluatie stelt op basis van deze gegevens een rapportage samen, die voor de planningsgroep als leidraad dient om het blok bij te stellen. Het oordeel van de studenten is van groot belang voor de inrichting van het onderwijs, zowel met betrekking tot de inhoud, de diepgang, als de wijze van aanbieden.

De rapportage van het project Programma-Evaluatie bevat een overzicht van het gemiddelde oordeel van alle studenten op de items uit de vragenlijst. Verondersteld wordt dat er een relatie bestaat tussen het oordeel van studenten en hun prestaties op de bloktoets. Dit is evenwel tot nu toe niet onderzocht. Een eventuele relatie zou van belang kunnen zijn voor het inrichten van het onderwijs door de planningsgroep. Bij de leden van de planningsgroep is bijvoorbeeld bekend in welke blokactiviteiten (practica, lezingen, trainingen) vooral basale leerstof wordt aangeboden, bij welke activiteiten dieper op de leerstof wordt ingegaan op basis van reeds aanwezig veronderstelde basiskennis (grotere diepgang), en waar de leerstof sterk schematisch aan bod komt. Tevens is bekend welke practica veel herhalingsaspecten bevatten. Elk van deze blokactiviteiten wordt door de studenten beoordeeld aan de hand van de vragenlijst programma-evaluatie door middel van de zogenaamde blokspecifieke items. Het is mogelijk dat de studenten die op de bloktoets relatief hoog scoren een ander oordeel hebben over blokactiviteiten met een hoog gehalte aan schematisch gebrachte basisstof en herhalingsaspecten dan studenten die relatief laag scoren op de bloktoets. Uiteraard kan dit ook gelden voor activiteiten waarvan de student alleen optimaal profijt kan hebben als de basisstof bekend is.

Elk jaar moet de planningsgroep weer een 'kosten-baten' afweging maken met betrekking tot de mate waarin basisleerstof, diepgang en herhaling moeten worden ingebouwd in de diverse activiteiten. Herhaling moet worden afgewogen tegen de tijdsinvestering van docenten en de beschikbaarheid van ruimte en materialen voor studenten. Met name op basis van een eventueel verschil in oordeel op de blokspecifieke items kan een planningsgroep besluiten tot bijstellen ten gunste van laag scorende studenten of hoog scorende studenten.

Om inzicht te verkrijgen in een mogelijke relatie tussen de oordelen van de studenten over de programma-onderdelen en hun prestaties op de bloktoets is een case-study uitgevoerd in blok 2.2, Bewegen en belasten, van het academiejaar 1992/1993. Het betreft een blok waarin vooral fysisch-diagnostische aspecten, inclusief functioneel-anatomische aspecten van het bewegingsapparaat en inspanningsfysiologie, centraal staan.

## Methode

Als indicator voor studieprestaties is uitgegaan van de scores op bloktoets 2.2. Op basis van deze scores zijn vijftien studenten geselecteerd met de hoogste scores (groep H) en vijftien studenten met de laagste scores (groep L). Bij deze selectie waren uitsluitend studenten betrokken die aan het einde van blok 2.2 de vragenlijst programma-evaluatie hebben ingevuld. Opgemerkt werd dat het percentage studenten dat de programma-evaluatie had ingevuld onder de laag-scorende studenten aanmerkelijk lager lag dan bij de hoog-scorenden.

De waardering voor het programma is gemeten aan de hand van de items in de programma-evaluatie. Deze vragenlijst kent een onderscheid in algemene (voor elk blok identieke) en blokspecifieke vragen. Om inzicht te kunnen krijgen in verschillen tussen de beoordelingen van groep H en L zijn de vragen op inhoudelijke gronden gebundeld in clusters.

De algemene vragen zijn samengebracht in de volgende clusters:

- Cluster 1: zaken betreffende voorkennis, doelstelling en studiegenot/beleving/interesse
- Cluster 2: functioneren van de onderwijs-groepen
- Cluster 3: trainingen in het skillslab en colleges
- Cluster 4: algemene waardering voor de organisatie en de leerzaamheid van het onderwijs.

Daarnaast zijn de blokspecifieke vragen in de volgende clusters samengebracht:

- Cluster 5: aansluiting van de skillslabtrainingen op de anatomiepractica en de deskundigheid van de begeleiding
- Cluster 6: begeleiding en leerzaamheid van de practica
- Cluster 7: presentatie en leerzaamheid van de colleges en de leerzaamheid van een computerprogramma.

Voor elk van de onderscheiden clusters is, zowel voor groep H als voor groep L, een gemiddelde clusterscore berekend. De vragen in cluster 1 tot en met 5 zijn oorspronkelijk op een vijfpuntsschaal beoordeeld (student is het 'volledig oneens' = 1, tot 'volledig eens' = 5 met een positief geformuleerde item), de vragen in cluster 6 en 7 op een tienpuntsschaal (een 'schoolcijfer' 1 tot en met 10). Om vergelijking mogelijk te maken zijn alle gemiddelde clusterscores van groep H en L getransformeerd naar een tienpuntsschaal.

## Resultaten

Achtereenvolgens zal worden ingegaan op de waardering (inclusief het verschil tussen de gemiddelde oordelen van groep H en L) voor:

- het programma, onderscheiden in de genoemde clusters (tabel 1);
- de individuele vragen waarvoor het verschil tussen groep L en H ten minste één schaalpunt bedroeg (tabel 2);

- met elkaar samenhangende reeksen van blok-activiteiten waarin onder meer herhalingsaspecten aan bod komen (tabel 3);
- de leerzaamheid van de colleges (tabel 4).

Om na te gaan of er sprake is van een statistisch significant effect, is een t-toets op de gemiddelde scores uitgevoerd. Indien het verschil tussen de beide groepen statistisch significant is ( $p < .05$ ), dan is dit aangegeven door een ster.

Uit tabel 1 blijkt dat groep H voor alle clusters hogere scores toekent dan groep L. De verschillen zijn het kleinst voor:

- cluster 4: algemene waardering voor de organisatie en leerzaamheid van het onderwijs;
- cluster 6: begeleiding en leerzaamheid van de practica;
- cluster 7: presentatie en leerzaamheid van de colleges.

De verschillen zijn het grootst ( $>1$ ) voor:

- cluster 3: trainingen in skillslab en colleges;
- cluster 5: aansluiting van de skillslabtrainingen op de anatomiepractica en de deskundigheid van de begeleiding.

Uit tabel 2 blijkt dat in het algemeen geldt dat groep H aangeeft meer geleerd te hebben in dit blok in vergelijking met groep L. Deze studenten lijken de inhoud van het blok interessanter te vinden en tevens wordt voor de deskundigheid van de begeleiding een hogere waardering gegeven. Opgemerkt dient te worden dat er geen vragen zijn waarbij groep L meer dan één schaalpunt hoger scoorde dan groep H. Over het geheel genomen waarderen de studenten uit groep H het computerprogramma ELBOW aanmerkelijk hoger dan de studenten uit groep L. Hierbij dient opgemerkt te worden dat er binnen de groepen een grote spreiding is wat betreft de scores.

## Practica

Het blok bevat een aantal practica met een sterk herhalingseffect. Dezelfde studiestof komt vanuit verschillende invalshoeken aan bod. Voor zowel de regio schouder/arm als voor de

regio nek/rug wordt een reeks opeenvolgende activiteiten aangeboden, respectievelijk een introductiecollege, een macroscopisch-anatomisch practicum, een anatomie in vivo practicum en het fysisch-diagnostisch onderzoek (skillslab training). De gemiddelde beoordeling voor de leerzaamheid van deze twee reeksen is voor groep H en groep L berekend. De gemiddelde scores staan vermeld in tabel 3.

Het kleinste verschil tussen groep H en L is te vinden bij het practicum anatomie in vivo. Dit type practicum bevat veel herhaling, vooral van het macroscopisch-anatomisch practicum en, in mindere mate, van het introductiecollege. Het anatomie in vivo practicum wordt in beide reeksen door beide groepen het laagst beoordeeld. Voor de reeks schouder/arm geldt, dat het verschil tussen de beoordeling van het macroscopisch-anatomisch practicum en het anatomie in vivo practicum in groep H groter is ( $8.10 - 6.50 = 1.60$ ) dan in groep L ( $7.33 - 6.70 = 0.63$ ). Voor de reeks rug geldt dit in mindere mate (verschil groep H;  $7.93 - 7.25 = 0.68$ , verschil groep L;  $7.20 - 7.00 = 0.20$ ).

Het macroscopisch-anatomisch practicum en de skillslabtraining worden door groep H hoger beoordeeld dan door groep L.

## Colleges

Niet alle colleges in dit blok zijn identiek wat betreft de diepgang, de hoeveelheid basisstof, de mate waarin de docent de stof geschematiseerd behandelt etcetera. Voor groep H en L is de gemiddelde waardering voor de leerzaamheid en presentatie van de colleges berekend (tabel 4).

Uit tabel 4 blijkt dat groep H de colleges A, B en C leerzamer vindt dan groep L. Voor het college D is er nauwelijks verschil, en voor E geldt het tegenovergestelde. Wanneer we naar de opzet van de laatste twee colleges kijken kan geconstateerd worden dat deze, in vergelijking met de colleges A, B en C, meer schematisch gepresenteerd worden (schema's op het bord etcetera) en dat hierin met name de basisstof

**Tabel 1.** Overzicht van de gemiddelde scores voor de clusters 1 tot en met 7 onderscheiden naar groep H (N=15) en L (N=15) en de verschillscore (H-L)

	Cluster						
	1	2	3	4	5	6	7
Groep H	8.02	7.49	8.44	7.37	8.28	7.39	7.19
Groep L	7.04	6.80	7.42	6.93	7.21	7.19	6.72
Verschilscore	0.98*	0.69	1.02*	0.44*	1.07*	0.20	0.47

**Tabel 2.** Overzicht van de items waarvoor groep H meer dan één schaalpunt hoger scoorde dan groep L

behorend tot cluster	verschilscore (H-L)
(1) Ik heb in dit blok veel opgestoken	1.73*
(1) Ik vond de leerstof interessant	1.33*
(2) De taken waren duidelijk omschreven	1.07
(3) Skillslabtrainingen sloten goed aan bij het blokprogramma	1.07*
(3) Skillslabtrainingen waren leerzaam	1.20*
(5) Aansluiten training schouder op anatomie	1.15*
(5) Deskundige begeleiding onderzoek schouder	1.14*
(5) Deskundige begeleiding onderzoek elleboog	1.37*
(5) Deskundige begeleiding onderzoek rug	1.62*
(5) Deskundige begeleiding onderzoek pols/hand	1.64
(7) Leerzaamheid computersimulatie programma ELBOW	1.72

**Tabel 3.** Overzicht van de gemiddelde score voor de leerzaamheid van het introductiecollege, de anatomiepractica en de skillslabtraining onderscheiden naar de regiones schouder/arm en rug

Practica	Schouder/arm			Rug		
	H	L	H-L	H	L	H-L
Introductiecollege	7.50	6.92	0.58	7.69	7.10	0.59*
Macroscopie practicum	8.10	7.33	0.77*	7.93	7.20	0.73*
Anatomie in vivo	6.50	6.70	-0.20*	7.25	7.00	0.25
Skillslabtraining	8.40	7.20	1.20	8.40	7.20	1.20

**Tabel 4.** Overzicht van de gemiddelde score voor de leerzaamheid van de colleges onderscheiden voor groep H en groep L

Colleges	H	L	H-L
A. Schouder/arm	7.50	6.92	0.58
B. Wervelkolom	7.69	7.10	0.59
C. Controle van de beweging	7.64	7.21	0.43
D. Biomechanica	6.55	6.67	-0.12
E. Spierfysiologie	6.86	7.23	-0.37

wordt uitgelegd (deels zelfs middelbare school leerstof). In de colleges A, B en C worden ook inzichten aangedragen die niet direct in een standaard anatomieboek terug te vinden zijn.

Mogelijk vormt dit een verklaring voor de hogere waardering voor de colleges A, B en C bij de studenten die hoog scoorden op de bloktoets en voor het ontbreken van verschil tussen hoog en laag scorenden (of een negatief verschil) met betrekking tot de colleges D en E. De verschillen tussen beide groepen zijn echter statistisch niet-significant. Deze resultaten moeten dan ook met de nodige voorzichtigheid geïnterpreteerd worden en opgevat worden als een mogelijke indicatie van de oorzaak van de gevonden verschillen in waardering.

## Discussie

Het blijkt dat in de groep die laag scoort op de bloktoets meer studenten voorkomen die geen vragenlijst Programma Evaluatie invullen, dan in de groep die hoog scoort. Mogelijk is dit een indicatie dat studenten die laag scoren op de toets minder bij het onderwijs betrokken zijn dan de studenten die hoog scoren op de bloktoets. In deze case-study zijn de vragenlijsten tijdens de laatste onderwijsgroepsbijeenkomst ingevuld, dat wil zeggen vóórdat de bloktoets werd afgenomen. Het is daardoor uitgesloten dat een (verondersteld) slecht of goed resultaat op de bloktoets van invloed is geweest bij het invullen van de vragenlijst.

In deze case-study zijn uitsluitend de oordelen betrokken van twee groepen: studenten die op de bloktoets behoren tot de extreem laag- of hoog-scorenden. De vraag is in hoeverre deze studenten ook bij andere toetsen tot de laag- of hogescorenden behoren. Om hiervan een indruk te krijgen, is van elk van de geselecteerde studenten het resultaat nagegaan op de voorafgaande bloktoets (blok 2.1) en de voortgangstoetsen van september en december 1992. Hiertoe is per toets de gehele populatie studenten gerangschikt naar score en vervolgens is een indeling in kwartielen (Q1, Q2, Q3,

Q4) gemaakt. Het kwartiel Q1 bestaat uit de 25% laagst scorende studenten; Q4 bevat de 25% hoogst scorenden. Nagegaan is hoeveel van de in groep H en L van de case-study geplaatste studenten in elk van deze vier kwartielen voorkomen. In tabel 5 staat het resultaat vermeld.

Uit tabel 5 kan afgeleid worden dat de studenten uit groep L voor bloktoets 2.1 praktisch allemaal onder de mediaan scoorden en de studenten uit groep H overwegend boven de mediaan. Voor de voortgangstoetsen geldt dit echter niet. Een eenduidige verklaring hiervoor ontbreekt op dit moment. Uitgezocht wordt of dit ook voor andere groepen studenten en andere toetsen geldt. Men kan aan de mogelijkheid denken dat de voortgangstoets een ander soort kennis meet dan de bloktoets.

Tussen groep L en H is sprake van een verschil in waardering voor het programma. Op bijna alle vragen van de programma-evaluatie geven studenten die een goed resultaat hebben behaald op de bloktoets een hogere waardering dan de studenten die een relatief lage score hebben behaald op de bloktoets. Het verschil is dus in ieder geval systematisch. Voor het optreden van deze verschillen zijn verschillende verklaringen denkbaar. Mogelijk hebben de laag-scorende studenten een onverschilligere/negatievere houding ten opzichte van de studieleiding; ze geven makkelijker lage beoordelingen voor het programma. De hoog-scorenden daarentegen zouden 'brave' studenten kunnen zijn, die 'lief' willen zijn voor de studieleiding. De verschillen zijn echter niet voor alle vragen even groot. Het verschil is het grootst (in ieder geval groter dan één schaalpunt op een tienpuntsschaal) bij vragen die betrekking hebben op het studiegenot, de waardering voor deskundige begeleiding bij het skillslab en de leerzaamheid van het computerprogramma ELBOW. Deze hogere beoordeling kan verband houden met een grotere tevredenheid over de betreffende activiteiten. Opmerkelijk is dat het vooral activiteiten betreft die een adequate voorkennis veronderstel-

**Tabel 5.** Overzicht van de resultaten op bloktoets 2.1 en twee voortgangstoetsen, onderscheiden voor groep H en groep L, verdeeld in kwartielen

	Bloktoets 2.1		VT-1		VT-2	
	L	H	L	H	L	H
Q 1	6	0	2	3	4	5
Q 2	7	2	4	2	4	3
Q 3	1	3	4	2	5	3
Q 4	0	10	4	8	2	4
Ontbreekt	1		1			

len bij de studenten. Het is mogelijk dat de hoog-scorende studenten ten tijde van hun deelname aan de blokactiviteiten deze voorkennis al wel hadden, terwijl deze nog ontbrak bij de laag-scorenden.

Bij het eerste contact (tijdens de colleges) is er tussen de groepen slechts een klein verschil in waardering. Het kleine, maar systematische verschil tussen de waardering voor de colleges heeft wellicht iets te maken met het gestructureerd brengen van basisstof (de spierfysiologie en biomechanica colleges) tegenover waardering voor inzichten die de basisstof ontstijgen (de colleges schouder/arm, wervelkolom en controle van beweging).

Herhalingsaspecten in practica worden meer gewaardeerd door studenten die laag scoren op een bloktoets dan door studenten die hoog scoren. In ieder geval zijn de verschillen tussen de waardering voor een macroscopisch-anatomisch practicum en een daaropvolgend anatomie in vivo practicum - met een hoog herhalingsgehalte - bij studenten die hoog scoren op de bloktoets groter dan bij de laag scorenden. Een verklaring zou kunnen zijn dat de hoog-scorenden de leerstof na één practicum reeds beter beheersen dan de laag-scorenden, met als gevolg dat zij herhaling van de leerstof overbodig vinden. Een andere verklaring zou kunnen zijn, dat de laag scorenden geen kans zien om optimaal gebruik te maken van de voorafgaande activiteiten. Ook is het denkbaar dat studenten zich niet voldoende voorbereiden op deze activiteiten. Met name

de aansluiting tussen het macroscopisch-anatomisch practicum schouder en de skillslab-training schouder werd door de hoog-scorende studenten meer gewaardeerd dan door de laag-scorenden. Het kan zijn dat deze laatsten de aansluitingsaspecten niet zien tengevolge van een slechtere anatomiekennis van de betreffende regio voorafgaand aan de skillslabtraining. Het verschil was voor de aansluiting van het macroscopisch-anatomisch practicum rug en de skillslabtraining rug minder uitgesproken. Het verschil tussen de beide regiones zit ondermeer in de hoeveelheid topografische anatomie-details die nodig zijn voor het optimaal verrichten van een onderzoek. De anatomie van de schouder/arm wordt meestal als ingewikkelder ervaren dan die van de rug. Ook een mechanica-simulatieprogramma (ELBOW) wordt door studenten die een beter resultaat op de bloktoets behalen gemiddeld genomen hoger gewaardeerd. Misschien zijn zij beter in staat of vinden zij het leuker om daadwerkelijk met de leerstof te 'spelen'. Opgemerkt dient te worden dat er binnen de groepen een groot verschil in waardering bestaat.

Een laatste mogelijkheid is dat er tussen de beide groepen een verschil in studiestijl bestaat. Het is denkbaar dat de laag-scorenden de aangeboden stof oppervlakkiger bestuderen, terwijl de hoog-scorenden, vanaf het eerste contact, het aangeboden studiemateriaal met een grotere diepgang (vaak uitgaande van al bij hen aanwezige basiskennis) tot zich nemen.

Overigens is van alle studenten de vooropleiding nagegaan. Tussen de beide groepen bestonden geen opmerkelijke verschillen.

### Ten slotte

Deze case-study is slechts kleinschalig opgezet. De vraag is in hoeverre deze verschillen ook binnen andere blokken gevonden zouden worden. Indien zou blijken dat deze casuïstische bevinding een algemeen karakter heeft, dan geldt dat studenten die hoog scoren op een bloktoets het studieprogramma anders waarderen dan zij die laag scoren. Een consequentie hiervan kan zijn dat de planningsgroep bij bijstelling van het blok een keuze moet maken voor aanbieden van meer gestructureerde (andere?) en op de basisstof gerichte studieprogramma's. Indien dit niet mogelijk is, moet voor elk onderdeel waarbij blijkt dat laag-scorenden ontevredener zijn dan hoog-scorenden in ieder geval duidelijk worden aangegeven dat een bepaalde activiteit pas leerzaam is als de basisstof bekend is. Om te voorkomen dat studenten onderwijsactiviteiten met ontoereikende voorkennis volgen, verdient het aanbe-

veling individuele leerwegen van studenten te stroomlijnen.

### Literatuur

1. Moust J, Bouhuijs P, Schmidt H. Probleemgestuurd leren. Hoger Onderwijs Reeks. Groningen: Wolters-Noordhoff, 1989.
2. Planningsgroep. Blokboek 2.2 "Bewegen en Belasten". Rijksuniversiteit Limburg, Faculteit der Geneeskunde. Academiejaar 1992/1993.

### DE AUTEURS

*H. van Mameren is als anatoom verbonden aan de vakgroep Anatomie/Embryologie van de Faculteit der Geneeskunde van de Rijksuniversiteit Limburg.*

*H.A.P. Wolfhagen is als onderwijskundige verbonden aan de vakgroep Onderwijsontwikkeling en Onderwijsresearch van de Rijksuniversiteit Limburg. Haar aandachtsgebied is de programma-evaluatie van het onderwijs binnen de Faculteit der Geneeskunde.*

*D. Riksen is als onderzoeksassistente verbonden aan de vakgroep Onderwijsontwikkeling en Onderwijsresearch van de Rijksuniversiteit Limburg.*

### Correspondentie-adres:

*H. van Mameren, Vakgroep Anatomie en Embryologie, Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht*