

Naar een interfacultaire voortgangstoets

Resultaten van proefafnames

C.P.M. van der Vleuten, G.M. Verwijnen, R.S.G. Holdrinet, B. Oeseburg, J.A. Bulte,
R.R.M. Leunissen

Samenvatting

Het gezamenlijk gebruik van toetsen is een zeer ongewone zaak in het onderwijs. Het gebruik van gezamenlijk toetsmateriaal biedt naast een kostenbesparing de mogelijkheid tot kwaliteitsbewaking en programma-evaluatie. Als proef, met het uiteindelijke doel te komen tot een structurele vorm van samenwerking op het gebied van voortgangstoetsen, zijn in het studiejaar 1994/1995 een tweetal Maastrichtse voortgangstoetsen afgenomen in Nijmegen. In december 1994 en maart 1995 werd de reguliere Maastrichtse voortgangstoets tegelijkertijd afgenomen bij de studenten in Maastricht en bij een deel van de studenten uit alle studie-jaren in Nijmegen. In deze bijdrage zullen enkele resultaten worden gepresenteerd. Hoewel de nodige voorzichtigheid geboden is bij de interpretatie van de resultaten als gevolg van de verschillende afnamecondities, komt naar voren dat de overeenkomsten in studieresultaat tussen beide faculteiten veel groter zijn dan de verschillen. Curriculumverschillen zijn nauwelijks in specifieke deelgebieden van de toets terug te vinden, maar komen pas op het niveau van individuele items naar voren. Naarmate het curriculum vordert, nemen ook deze verschillen af.

Inleiding

De productie van goede toetsen is een kostbare aangelegenheid. Het zou voor de hand liggen

dat faculteiten meer van elkaars werk op dit gebied profiteren. Het gebruik van gezamenlijk toetsmateriaal biedt naast kostenbesparing de mogelijkheid tot kwaliteitsbewaking en programma-evaluatie. Ondanks al deze voordelen is het gezamenlijk gebruik van toetsen ongebruikelijk.

Met de introductie van het nieuwe curriculum geneeskunde in Nijmegen werd de wenselijkheid vastgesteld van een aangepast toets-systeem, waarin ook een plaats voor een interdisciplinaire en integratieve toets noodzakelijk zou zijn. In plaats van zelf een dergelijk toetsontwerp te ontwikkelen werd de mogelijkheid nagegaan van het gebruik van de Maastrichtse voortgangstoets.¹ Bij wijze van proef zijn in het studiejaar 1994/1995 een tweetal Maastrichtse voortgangstoetsen afgenomen in Nijmegen. Inmiddels zijn plannen gemaakt voor een meer structurele vorm van samenwerking op korte en lange termijn. Vanaf het studiejaar 1995/1996 zal driemaal per jaar de Maastrichtse voortgangstoets in het nieuwe Nijmeegse curriculum worden gebruikt (dus in 1995/1996 alleen voor de propedeuse-studenten). Het betreft een qua tijd en inhoud volledig parallelle afname: tegelijkertijd wordt dezelfde toets op beide locaties afgenomen. Ook in Nijmegen heeft de toets consequenties voor de studievoortgang van studenten. In deze unilaterale vorm van samenwerking verstrekt Maastricht de kant-en-klare toets aan Nijmegen tegen een financiële vergoeding. Voor de langere termijn wordt een bilaterale samenwerking na-

gestreefd, waarin beide faculteiten materiaal leveren voor de toets.

De proefafnames in het studiejaar 1994/1995 waren onder meer bedoeld om de logistieke consequenties in kaart te brengen. Het bleek organisatorisch goed mogelijk te zijn om de voortgangstoets parallel af te nemen. In deze bijdrage zullen de studieprestaties van de studenten van beide faculteiten worden vergeleken. Naast resultaten op de gehele toets worden resultaten ook uitgesplitst naar onderdelen van de toets.

Methode

De parallelle afnames vonden plaats in december 1994 en in maart 1995. Voor Maastricht waren dit de reguliere toetsafnames. In Nijmegen werden studenten gerecruteerd uit alle studiejaren, maar op basis van vrijwilligheid. De toets heeft in Maastricht consequenties voor de studievoortgang, de toets had in Nijmegen geen enkele consequentie voor de studenten. Beide aspecten zijn wezenlijke methodologische tekortkomingen die een goede vergelijking van het kennisniveau bemoeilijken: vrijwillige deelnemers zijn meestal de betere studenten.² Toetsen die niet meetellen kunnen aanleiding geven tot andere invulstrategieën en met name het gokgedrag beïnvloeden. In tabel 1 zijn de deelnemende aantallen studenten weergegeven. De belangstelling voor deelna-

me in Nijmegen was voor de maarttoets beduidend lager. De selectie-bias kan hierdoor nog ernstiger zijn.

De decembertoets bestond uit 242 juist/onjuist/vraagteken-items en de maarttoets uit 226 items. Voor elke student werden goed-min-foutscores in procenten berekend. Om de invulstrategieën te vergelijken, werden de goed-, fout- en vraagtekenscores vergeleken.

Resultaten

Uit de vergelijking van de goed-, fout- en vraagtekenscores kwam een systematisch verschil naar voren. Ter besparing van ruimte worden niet al deze gegevens gerapporteerd. Echter, in alle jaargroepen werd door de Nijmeegse studenten een significant hogere goed- en foutscore, en een significant lagere vraagtekenscore verkregen (one-way Anova, $p < 0.001$). Met andere woorden, de Nijmeegse studenten vulden meer vragen in, maakten daardoor meer vragen goed, maar ook meer fout. Blijkbaar werd er, gemiddeld genomen, meer gegokt. Ook bleken de standaarddeviaties van scores binnen jaargroepen systematisch groter voor de Nijmeegse groep. Deze gegevens duiden erop dat de Nijmeegse studenten een andere invulstrategie hebben toegepast. De goed-min-foutscore, waarin voor gokken wordt gecorrigeerd, is dus de meest aangewezen score ter vergelijking, maar ver-

Tabel 1. Aantal deelnemende studenten aan de voortgangstoetsafnames

Jaar	December 1994		Maart 1995	
	Maastricht	Nijmegen	Maastricht	Nijmegen
1	190	124	218	54
2	146	104	151	60
3	135	87	133	51
4	188	151	169	53
5	144	140	155	69
6	135	122	140	74

Tabel 2. Gemiddelde goed-min-foutscores (Gem.) en standaarddeviaties (Sd.) voor de totale toets en clusters van disciplines, apart voor Maastricht en Nijmegen

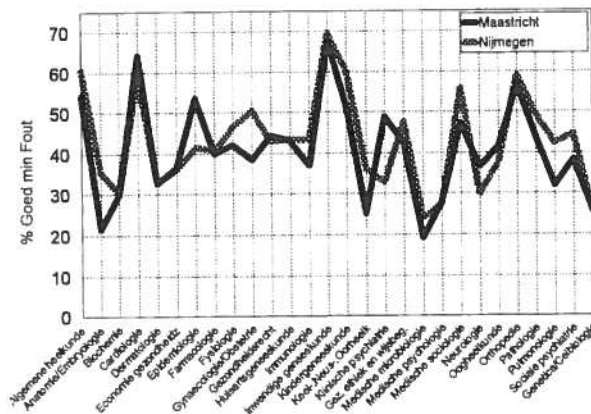
Jaar	December 1994				Maart 1995			
	Maastricht		Nijmegen		Maastricht		Nijmegen	
	Gem.	Sd.	Gem.	Sd.	Gem.	Sd.	Gem.	Sd.
<i>Totale toets</i>								
1	9.61	4.06	10.70	5.37	11.42	4.29	12.17	6.85
2	20.68	5.94	19.27	6.68	20.64	6.43	22.26	8.27
3	26.18	6.68	23.50	9.32	26.51	6.88	30.37 [†]	9.14
4	33.27	7.98	31.54	8.95	35.14	7.64	38.03	9.28
5	38.46	8.53	38.80	9.16	38.41	8.59	41.52	9.37
6	41.03	8.01	44.49 [†]	8.99	42.12	6.71	44.85	11.07
<i>Basisvakken</i>								
1	8.43	4.77	10.68 [†]	6.78	11.26	5.66	14.97 [†]	8.20
2	19.09	8.27	16.66	8.59	18.78	8.19	25.48 [†]	10.74
3	21.08	8.56	19.63	10.00	24.46	8.88	31.58 [†]	11.06
4	25.86	9.31	25.66	10.50	32.46	8.96	37.74 [†]	10.02
5	30.13	10.97	31.16	11.70	36.25	10.39	38.41	10.30
6	32.33	10.21	37.07 [†]	11.00	36.49	8.24	41.08 [†]	12.09
<i>Klinische vakken</i>								
1	8.24	4.74	9.23	6.65	4.77	4.57	3.74	7.90
2	19.26	6.94	20.46	8.13	14.41	7.46	13.33	8.53
3	26.39	8.62	26.09	10.60	23.10	8.69	25.02	10.99
4	37.88	9.80	36.15	10.80	34.38	10.33	35.41	11.92
5	44.56	9.71	45.30	10.40	38.72	11.06	41.77	13.35
6	48.71	10.04	52.51 [†]	11.10	46.44	9.12	48.60	12.85
<i>Gedragwetenschappelijke vakken</i>								
1	14.93	9.74	13.98 [†]	10.30	26.18	11.90	24.21	14.33
2	26.90	11.20	21.77 [†]	10.40	38.25	12.40	34.44	12.78
3	35.70	12.71	25.39 [†]	13.80	38.42	12.67	39.26	12.57
4	37.65	12.86	32.93 [†]	13.90	42.72	12.25	44.34 [†]	14.42
5	41.43	13.04	39.53	12.10	42.53	13.40	47.83 [†]	13.19
6	41.26	11.36	41.42	12.37	45.24	11.15	45.05	16.74

[†] p < 0.01 vergeleken met Maastrichtse gemiddelde (one-way ANOVA)

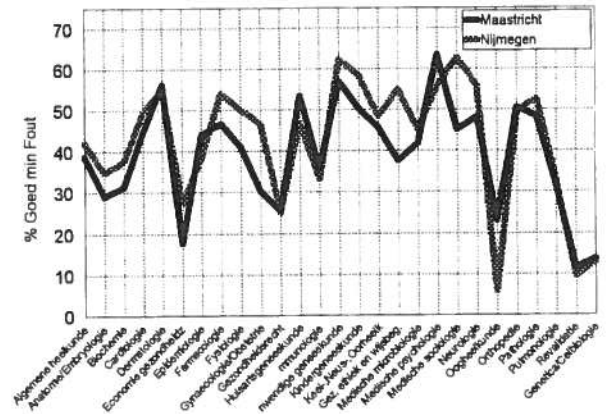
schillen in antwoordstrategie kunnen de gegevens beduidend beïnvloeden.

In tabel 2 zijn de resultaten weergegeven van de goed-min-foutscores voor de totale toets en voor de drie clusters van vakgebieden. Ook voor de goed-min-foutscores blijkt wederom dat de standaarddeviaties van de Nij-

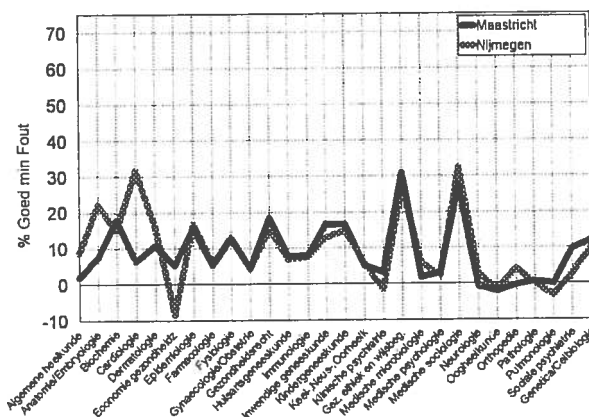
meegse groepen studenten systematisch groter zijn dan van de Maastrichtse studenten. Voorts zijn er een aantal statistisch significante verschillen, maar deze blijken niet systematisch voor beide metingen. Alleen voor de basisvakken op zesdejaarsniveau blijkt zowel voor de decembertoets als voor de maarttoets een sig-



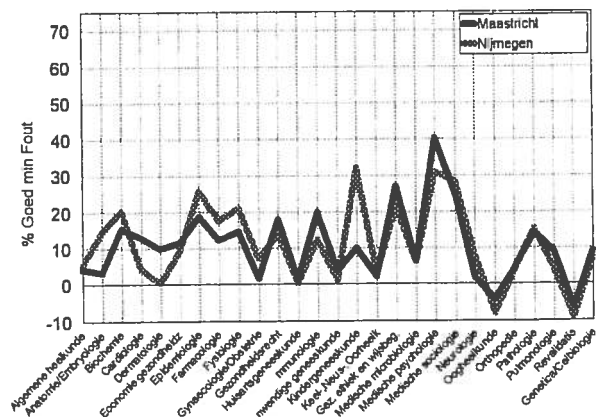
Figuur 1a. Vakprofiel 6e jaars december 1994



Figuur 1b. Vakprofiel 6e jaars maart 1995



Figuur 1c. Vakprofiel 1e jaars december 1994

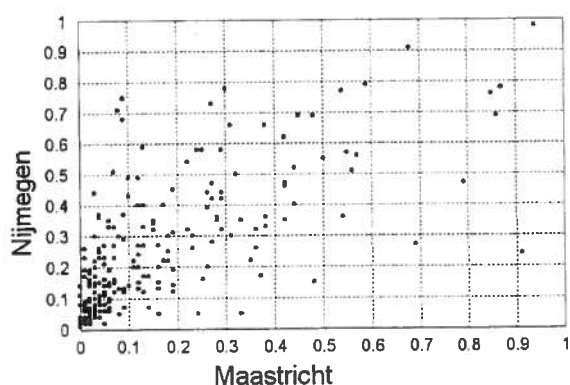


Figuur 1d. Vakprofiel 1e jaars maart 1995

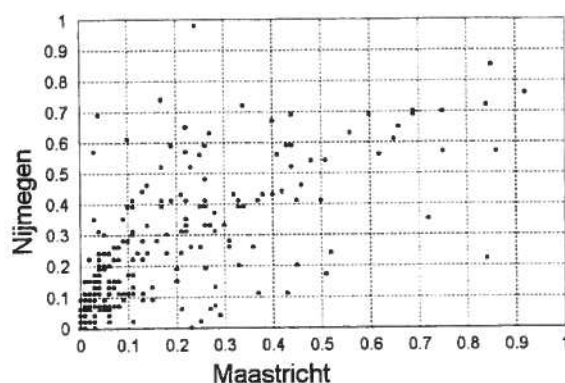
nificant verschil te bestaan. Opmerkelijk is echter dat dit verschil weer niet bestaat voor de vijfdejaars. Binnen één toetsmoment zijn er vooral significante verschillen voor de basisvakken in de maarttoets ten gunste van Nijmegen en voor de gedragswetenschappelijke vakken in de decembertoets in de doctoraalfase van de studie ten gunste van Maastricht. Opvallend is vooral de parallelliteit van de 'groei' in kennis voor beide faculteiten. Dit ondanks de verschillen in curricula en de verschillen in invulstrategie.

De mate van parallelliteit in kennisopbouw kan

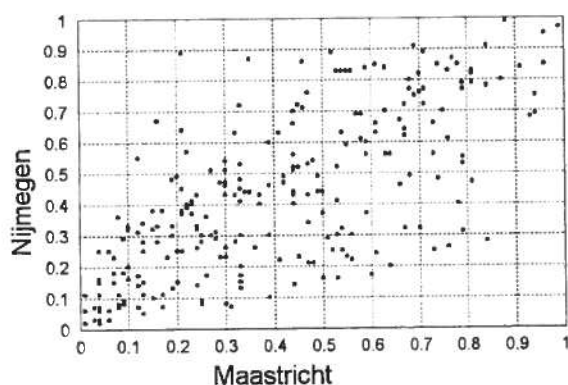
nader bestudeerd worden door een verdere uitsplitsing van de resultaten. In figuur 1a tot en met 1d is dat grafisch gedaan voor de verschillende vakgebieden in de voortgangstoets. In de figuren zijn de gemiddelde resultaten afgebeeld, apart voor de zesdejaars en de eerstejaars. Men moet zich daarbij realiseren dat aan de deelscores op vakgebieden slechts een klein aantal vragen ten grondslag ligt. De spreiding rondom de gemiddelde waarden is als gevolg daarvan erg groot. Significantietsing is om die reden achterwege gelaten. Meer van belang is echter de inspectie van het profiel. De mate



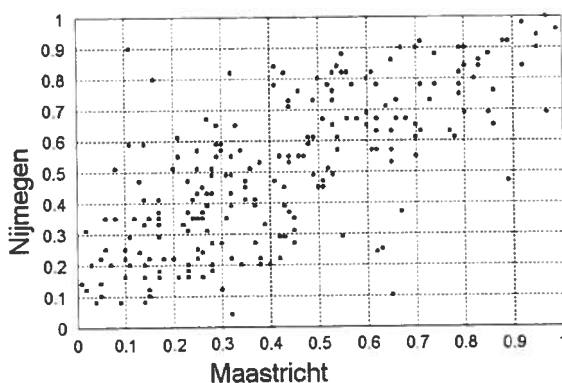
Figuur 2a. P-waarden jaar 6, december 94



Figuur 2b. P-waarden jaar 1, maart 95



Figuur 2c. P-waarden jaar 3, december 94

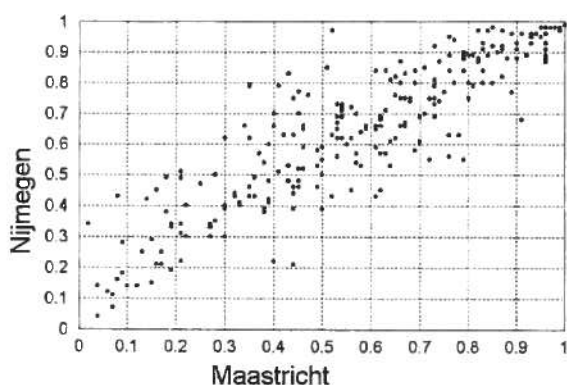


Figuur 2d. P-waarden jaar 3, maart 95

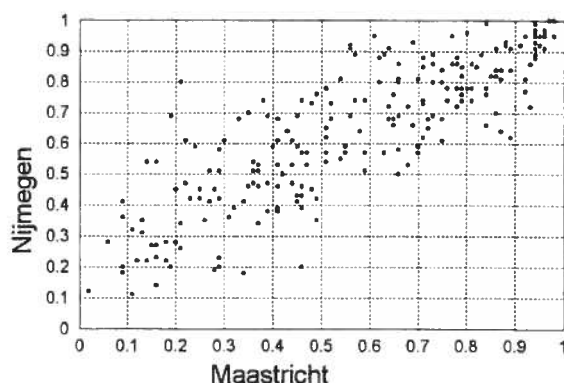
waarin de lijnen van beide faculteiten elkaar volgen is opmerkelijk. Een vakgebied is blijkbaar binnen een bepaalde toets voor een bepaalde jaargroep moeilijker of makkelijker ongeacht de facultaire origine. Het scoreprofiel op de categorieën van de blauwdruk (met name orgaancategorieën) is niet afgebeeld, maar levert een overeenkomstig beeld op.

Het meest gedetailleerde niveau van vergelijking betreft een vergelijking van itemscores. In figuur 2a tot en met 2f zijn de p-waarden (het percentage studenten dat de vraag goed beantwoord heeft) voor beide faculteiten in een scatterdiagram geplaatst voor een drietal jaargroepen. Naarmate de punten in de wolk verder van de diagonaal liggen zou gesproken kunnen worden van een specifiek effect van het onder-

wijsprogramma. De figuren 2a en 2b bevatten de resultaten van het eerste jaar. Een groot cluster van punten bevindt zich nabij de oorsprong (veel vragen waarover in beide faculteiten geen kennis bestaat). Ook zijn er items die van de oorsprong verwijderd zijn en die nabij de diagonaal liggen. Dit zijn de items waarover in beide faculteiten gelijke kennis bestaat. Daarnaast zijn er echter tamelijk veel items die buiten de diagonaal liggen en die hetzij in Nijmegen goed worden gemaakt hetzij in Maastricht. In het derde jaar (figuur 2c en 2d) zijn er heel wat minder items nabij de oorsprong en nog steeds is er sprake van een flinke bandbreedte in het scatterdiagram, duidend op curriculumspecifieke kennis op een ruim aantal items. In het zesde jaar (figuur 2e



Figuur 2e. P-waarden jaar 6, december 94



Figuur 2f. P-waarden jaar 6, maart 95

en 2f) is een groep items geclusterd nabij de andere hoek van het diagram: voor beide faculteiten zijn dit de items die (nagenoeg) volledig beheerst worden door de gehele groep studenten. De puntenwolk is nu echter beduidend minder breed en de meeste items liggen dichter bij de diagonaal. Blijkbaar wordt een vergelijkbare kennisopbouw verkregen aan het einde van het curriculum. Overigens vormen de niet afgebeelde jaren keurige tussenliggende puntenwolken in dezelfde lijn.

Discussie

De proefafnames laten een beeld zien dat sterk overeenkomt met eerdere uitkomsten van interfacultair vergelijkend onderzoek tussen medische faculteiten in Nederland: de overeenkomsten in uitkomsten dringen zich meer op dan de verschillen.³⁻⁹ Voorzover er sprake is van enig verschil, wordt dat voornamelijk gevonden in de basisvakken ten gunste van Nijmegen en in de gedragswetenschappelijke vakken ten gunste van Maastricht. De effecten zijn echter niet systematisch gevonden voor beide toetsen. Bovendien dient ernstig rekening te worden gehouden met fluctuaties in uitkomsten als gevolg van de verschillen in invulstrategie tussen de Maastrichtse en de Nijmeegse studenten. Wellicht doordat de toetsen geen consequenties hadden voor de Nijmeegse stu-

denten hebben zij gemiddeld meer gegokt bij het beantwoorden van de vragen dan de Maastrichtse studenten.

Deze studie laat echter ook zien dat de parallelliteit in uitkomsten zich ook uitstrekt tot resultaten op een nog gedetailleerder niveau. De profielen van scores op vakgebieden volgen een vergelijkbaar beeld voor beide faculteiten. De moeilijkheidsgraad van een groep vragen binnen een cluster zoals een vakgebied (maar ook blauwdrukcategorie) is blijkbaar een tamelijk universeel gegeven, dat onafhankelijk is van het gevolgde curriculum.

Curriculumverschillen bleken wel uit individuele vragen. In het eerste jaar worden vragen gevonden die door de ene dan wel de andere groep goed worden beantwoord. Dat effect neemt systematisch af in hogere studiejaren, dus naarmate het curriculum vordert. Aan het einde van het curriculum komt de parallelliteit ook op itemniveau tot uiting.

Literatuur

1. Vleuten CPM van der, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based curriculum. *Med Teacher*, ter perse.
2. Cate THJ ten. De invloed van anonimiteit op de resultaten van tijdschrijfonderzoek. *Tijdschrift voor Onderwijsresearch* 1985; 10: 263-73.

3. Bender W, Cohen-Schotanus J, Imbos Tj, Versfelt W, Verwijnen GM. Medische kennis bij studenten uit verschillende faculteiten: van hetzelfde laken een pak? Ned Tijdschr Geneesk 1984; 128: 917-21.
4. Verwijnen GM, Vleuten CPM van der, Imbos Tj. Comparing an innovative medical school with traditional schools: an output analysis in the cognitive domain. In: Nooman ZM, Schmidt HG, Ezzat ES, redacteurs. Innovation in medical education: an evaluation of its present status. New York: Springer Publishing Company, 1990: 40-9.
5. Cate ThJ ten. De algemene medische kennistoets: resultaten. Amsterdam: Universiteit van Amsterdam; 1986 Interne Publicatie NPOG-MEMO 86-I.
6. Bosschert D. Resultaten voortgangstoets 30 juni 1986. Leiden: Rijksuniversiteit Leiden; 1988 Intern rapport Sectie Onderwijsontwikkeling.
7. Imbos Tj, Drukker J, Mameren H van, Verwijnen M. The growth in knowledge of anatomy in a problem-based curriculum. In: Schmidt HG, Volder ML de, redacteurs. Tutorials in problem-based learning. Assen: Van Gorcum, 1984: 106-15.
8. Jaarverslag Evaluatie Project Geneeskunde. Maastricht: Rijksuniversiteit Limburg; 1995 Interne Publicatie EPG-publ.nr. 95-01: 53.
9. Hessen PAW van, Verwijnen GM, Imbos Tj. De kennis van de Nederlandse basisartsen gemeten met de Maastrichtse voortgangstoets. Ned Tijdschr Geneesk 1991; 135: 1975-8.

DE AUTEURS

C.P.M. van der Vleuten, psycholoog, is als universitair hoofddocent verbonden aan de vakgroep Onderwijsontwikkeling en Onderwijsresearch, Rijksuniversiteit Limburg.

G.M. Verwijnen, huisarts, is voorzitter van de Voortgangstoets Beoordelings Commissie van het Evaluatie Project Geneeskunde, Rijksuniversiteit Limburg.

R.S.G. Holdrinet, internist-haematoloog en universitair hoofddocent, is voorzitter van de Bestuurscommissie Herziening Curricula, belast met de herziening van het curriculum geneeskunde aan de Faculteit der Medische Wetenschappen van de KU-Nijmegen.

B. Oeseburg, hoogleraar fysiologie en vice-voorzitter van de Bestuurscommissie Herziening Curricula aan de Faculteit der Medische Wetenschappen van de KU-Nijmegen.

J.A. Bulte, onderwijskundige, senior-beleidsmedewerker in de staf Onderwijs- en Studentenzaken van de Faculteit der Medische Wetenschappen, KU-Nijmegen.

R.R.M. Leunissen, medisch doctorandus, hoofd onderwijscoördinatie in de staf Onderwijs- en Studentenzaken van de Faculteit der Medische Wetenschappen, KU-Nijmegen.

Correspondentie-adres:

Dr. C.P.M. van der Vleuten, Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Faculteit der Geneeskunde, Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.