

In 1972 begon Nederland met een uniek experiment dat weldra zal worden beëindigd: de toelating van student tot de opleiding tot arts via een systeem van gewogen loting.¹ Veertig jaar heeft dat systeem gefunctioneerd en meerdere generaties artsen zijn via deze route in de studie en het beroep terecht gekomen.

De ratio, de emotie, en het Hawthorne-effect

Selectie voor de geneeskunde opleiding

Prof dr Th.J. ten Cate

Expertisecentrum voor Onderwijs en Opleidingen UMC Utrecht²

Nederland heeft in deze decennia verbazing gewekt in het buitenland. In geen enkel ander weldenkend land worden studenten aan het begin van de artsopleiding via loting toegelaten. Waarom kon dit in Nederland, waarom heeft het systeem veertig jaar kunnen functioneren en wat is er gebeurd dat het roer nu om moet? En wat is het alternatief? Selectie op basis van door faculteiten veronderstelde criteria voor geschiktheid voor de medische opleiding gaat in de praktijk gepaard met een grote kans op foute beslissingen en is daarom niet 'eerlijker' dan gewogen loting. Maar er is wel iets anders aan de hand. Er zijn sterke aanwijzingen dat het selectieproces op zichzelf een vormende werking heeft. Selectie moet daarom eerder worden beschouwd als een interventie dan als een meetinstrument. Psychologen zouden dit een Hawthorne-effect noemen.

De ratio van de gewogen loting

Nederland heeft acht medische opleidingen die wel enigszins, maar niet zeer wezenlijk van elkaar verschillen. Alle opleidingen voldoen aan eindtermen die gezamenlijk zijn vastgesteld in een sinds 1994 regelmatig vernieuwd Raamplan Artsopleiding, gekoppeld aan de registratie tot arts en er is dus geen wezenlijke profilering die rechtvaardigt dat verschillende typen studenten worden aangetrokken. De uitstroom uit de artsopleiding is dus kwalitatief ongeveer gelijk in Nederland en er is een extern kwaliteitszorg- en visitatiesysteem dat dat bewaakt.

Die gelijkheid geldt ook voor de instroom. Studenten moeten voldoen aan vwo-pakketisen om geneeskunde te studeren en arts te worden. De medische opleidingen

hebben een goed studierendement; studenten zijn voor het overgrote deel goed gemotiveerd om arts te worden en bereid hiervoor de gevraagde inspanning te leveren. Van de circa 20 procent studenten die niet de eindstreep haalt, heeft een belangrijk deel een andere levenskeuze gemaakt of heeft een deel fysieke of mentale problemen. Een minderheid zou men als 'ongeschikt' kunnen aanmerken. Er zijn ook studenten die wel de eindstreep halen, maar door opleiding soms als ongeschikt worden beschouwd. De opleiding heeft weinig instrumenten om hen tegen te houden omdat zij wel studieresultaten behalen, maar zou wensen dat zij niet zouden zijn toegelaten. Dit percentage is niet precies bekend, maar wordt geschat op minder dan 5 procent.

Ook de Nederlands gezondheidszorg kent geen wezenlijk kwaliteitsprobleem waarvoor selectie van studenten van artsen een goede oplossing is. De gezondheidszorg is beter dan in veel andere landen, mede door haar beroepsbeoefenaren.

Er dus weinig rationele aanleiding om een kwalitatieve selectie toe te passen. De Nederlandse volksaard heet rationeel te zijn en zich niet al te veel aan te trekken van gewoontes in andere landen en dat is misschien de reden dat gewogen loting zo lang is gehandhaafd. Het systeem is efficiënt en doorzichtig en behandelt kandidaten min of meer gelijk. Het is door zwaargewichten in de psychometrie en arbeids- en organisatiepsychologie verdedigd.³

Dat er geen zuivere maar altijd gewogen loting plaatsvond is ook rationeel. Er bestaat immers wel een correlatieve verband tussen eindexamencijfers en studieresultaten, zij het beperkt en vooral vroeg in de opleiding.

	Afgewezen (60%)	Aangenomen (40%)
Geschiedt (80%)	48%	32%
Ongeschiedt (20%)	12%	8%

Tabel 1. Veel geschikte kandidaten moeten worden afgewezen.

De emotie rond de gewogen loting

In 1996 laaiden de emoties rond het systeem opeens hoog op toen een middelbare scholiere met hoge cijfers niet werd ingeloot. Het was wachten op zo'n incident, want ook bij gewogen loten blijft er een kans, hoe uitzonderlijk ook, dat studenten met hoge vwo-cijfers de loting niet passeren. De emoties liepen zo hoog op dat het systeem werd aangepast. Niet alleen werden alle studenten met gemiddelde eindexamencijfers van 8 en hoger voortaan toegelaten, de politiek besloot faculteiten te stimuleren om te gaan experimenteren met decentrale selectie voor een beperkt aantal plaatsen, aanvankelijk 10 procent, later 50 procent. Loting moest echter mogelijk blijven, in het licht van de gedachte dat ook studenten die minder hoge eindexamencijfers behalen enig recht behouden om geneeskunde te gaan studeren. Emoties bleken uit talloze ingezonden stukken in kranten, interviews met afgewezen kandidaten. Er bleek grote maatschappelijke weerstand te zijn ontstaan tegen de gewogen loting. Dat is begrijpelijk, vanuit een psychologisch en emotioneel gezichtspunt. De mens wil graag 'het lot in eigen hand' hebben, een slogan die regelmatig gebruikt is in de discussies over selectie en loting. Iedereen wil het gevoel hebben: *Als ik mijn best doe, kan ik bereiken wat ik wil. Als ik echt dokter wil worden, moet dat in een vrij land kunnen. En als ik het dan niet word, dan wil ik graag mezelf verwijten dat ik niet genoeg mijn best heb gedaan. Maar ik vind het onacceptabel als ik geen invloed kan uitoefenen op mijn lot, en dat een dobbelsteen dat bepaalt.* Niet alleen van aankomende studenten komt deze weerstand. Ook vanuit het medische beroep. Er heerst een overtuiging dat er betere dokters op de arbeidsmarkt komen als zij voor de poort van de artsopleiding worden geselecteerd. Het maatschappelijk klimaat waarin gewogen loting acceptabel was, is veranderd.

De beperkte ratio van een kwalitatieve selectie

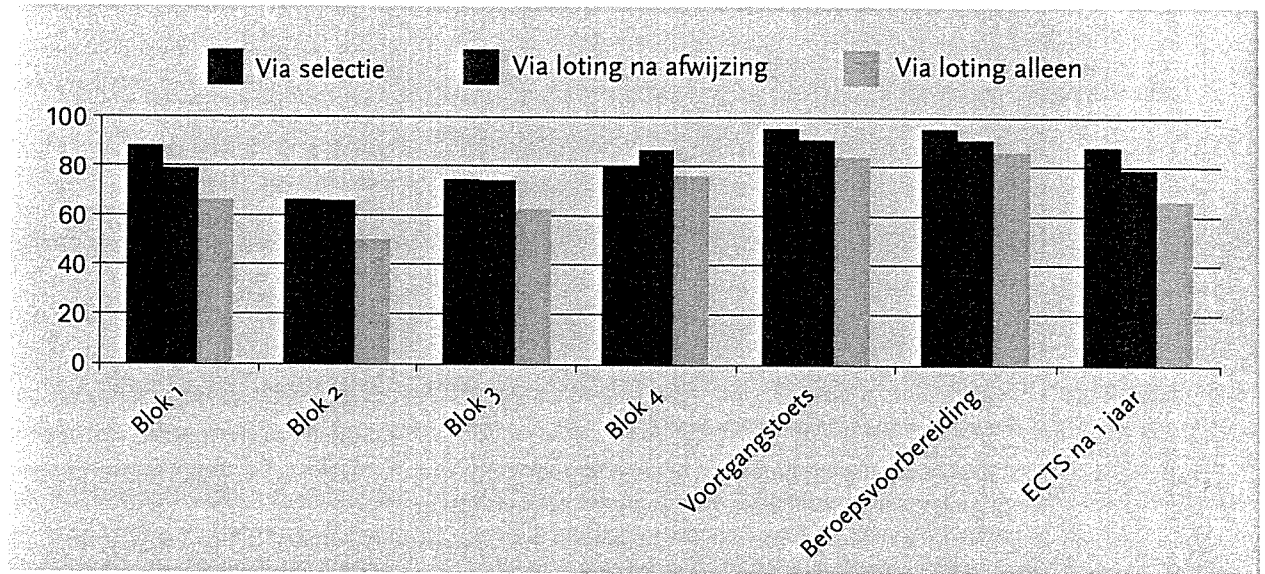
Het probleem van de toelatingsbeperking is dat er in het algemeen op de drie goede kandidaten slechts één kan worden toegelaten. Het is dus niet: 'Hoe zorgen we dat

alleen goede kandidaten worden toegelaten tot de opleiding?', maar: 'Hoe houden we geschikte kandidaten buiten de opleiding?'. Het antwoord van velen is: 'Niemand is gelijk en de besten moet worden gekozen'. Dat klinkt logisch maar is veel lastiger dan het lijkt. Daar zijn twee redenen voor: (1) er is zelden overeenstemming te vinden is over wat onder 'best' moet worden verstaan en (2) er is geen goed instrument is om dat 'beste' betrouwbaar te meten. Als de verschillen tussen de kandidaten klein zijn, een typisch Nederlandse situatie, is het meetprobleem groot.

Er zijn in de wereld veel gegevens verzameld over selectie voor de medisch opleiding, zowel over de betrouwbaarheid van procedures, c.q. over de mogelijkheid om een helder, liefst reproduceerbaar onderscheid tussen kandidaten te maken, als over de validiteit, c.q. over de vraag of de criteria waarop de kandidaten onderscheiden worden verdedigbaar zijn in het licht van de wensen voor het type beroepsbeoefenaren voor de gezondheidszorg of het type studenten dat de opleidingen graag zien. Ook in Nederland is en wordt onderzoek gedaan sinds decentrale selectie mogelijk is.

De overtuiging heerst dat er betere dokters komen als zij worden geselecteerd

Selectie op basis van cognitieve en niet cognitieve criteria wordt door medische opleidingen van belang geacht. Salvatori rapporteert in een literatuurstudie dat cognitieve criteria zoals cijfers van het secundair onderwijs of toelatingsexamens studieprestaties kunnen voorspellen, maar dat de correlaties niet erg hoog zijn, en ook dat niet-cognitieve selectiecriteria weinig betrouwbaar en valide zijn.⁴ Eva ontwikkelde in Canada een Multiple Mini-Interview procedure die momenteel geldt als de meest betrouwbare en valide selectiemethode in het medisch onderwijs. De kandidaat passeert circa 10 korte "stations" waarin een gesprek gevoerd of vaardigheid getoond wordt. In vergelijking met andere instrumenten correleerde deze MMI het best met de Medical Council of Canada Qualifying Examination part II vijf jaar later (0.35). Andere criteria (een autobiografie, Grade Point Average voorafgaand aan de opleiding, interviews en



Figuur 1. Slaagprestaties van geselecteerde en afgewezen studenten voor studieonderdelen in Groningen. Wat men eerder zou verwachten is het volgende gemiddeld zouden de prestaties van studenten in de blauwe en rode balken samen gelijk zijn aan die van de studenten van de groene balken (de 'alleen loting' groep) en die van de geselecteerden boven de groene balken uitstijgen en die van de afgewezenen achterblijven bij de groene balken. De selectie zou dan een mooi onderscheid hebben gemaakt binnen de groep die anders tot de loting zouden behoren. Dat is dus niet het geval. De afgewezenen doen het beter dan je zou verwachten.

een simulatie-tutorial) vertonen correlaties van 0.7 tot 0.14. Een correlatie van 0.35 klinkt niet gek, maar staat gelijk aan een verklaarde variantie van 12 procent (het kwadraat van de correlatie). Met andere woorden, ongeveer 90 procent van de variantie in het aantal behaalde MCCQE-II stations is niet verklaard door selectiecriteria aan de poort. Een beter criterium nog dan het aantal stations waarvoor men slaagde is de gemiddelde MCCQE-II score voor de stations. Die correleerde slechts 0.19 met de MMI.⁵ Deze publicatie geeft met 0.35 één van de hoogst beschreven correlaties van selectiecriteria met een *outcome* maat over langere termijn. De voorspellende waarde van selectie-instrumenten is niet hoog. Hierbij moet men verder bedenken dat deze correlaties uitsluitend zijn gebaseerd op kandidaten die door de selectie zijn heen gekomen. Het is zeer wel mogelijk dat veel afgewezen kandidaten toch geschikt zouden zijn. Tabel 1 is een realistisch voorbeeld voor de geneeskundeopleiding. Op basis van de ervaringen met gewogen loting in de afgelopen decennia is de veronderstelling gerechtvaardigd dat zeker 80 procent van de scholieren die zich aanmelden voor de geneeskundeopleiding in theorie geschikt zijn. Als 40 procent van hen kan worden geplaatst, zal 48 procent van de totale groep vallen in de categorie 'geschikt, maar afgewezen.'

Bij zeventuizend aanmeldingen voor circa drieduizend plaatsen gaat het om ruim drieduizend afwijzingen van - op zich - geschikte kandidaten. Dit getal kan lager zijn bij een weging in de loting of bij selectie, veel minder zal het niet zijn. Er is domweg geen ruimte is om alle

geschikte kandidaten toe te laten. Dat is pijnlijk en frustrerend, ongeacht of men selecteert of loot.

De kans is groot dat veel afgewezen kandidaten prima studieresultaten zouden hebben vertoond als zij waren toegelaten. Er is hier ook wat empirische evidentie voor. Prof Janke Cohen-Schotanus laat in haar oratie zien dat cognitieve studieresultaten van medische studenten die aanvankelijk door haar faculteit zijn afgewezen voor een kwalitatieve selectieprocedure (o.m. met MMI), maar daarna toch via loting zijn begonnen met de studie, in het eerste jaar ongeveer gelijk zijn aan geselecteerde studenten. De getallen in figuur 1 zijn afkomstig uit deze oratie.⁶

Selectieprocedures voor medische opleidingen kennen meestal meerdere onderdelen. Vaak is er een voorselectie en daarna een echte selectie met een kleinere groep. De selectie kan bestaan uit een kennistest, een gesprek met een commissie, een assessment, een opdracht, deelname aan een onderwijsactiviteit, een stationstoetsing zoals de MMI of andere methoden.

De meest betrouwbare en valide methoden zijn ook het meest kostbaar; vaak te kostbaar om voor alle kandidaten toe te passen. Daarom wordt bij veel belangstellenden de voorselectie op basis van aanmeldingsformulieren uitgevoerd. Die procedures zijn minder betrouwbaar en valide, zeker als het gaat om formulieren. Dat betekent dus dat opleidingen die intensieve procedures uitvoeren, zoals de MMI, om zo betrouwbaar en valide mogelijk te selecteren eerst veel kandidaten afwijzen op minder betrouwbare gronden.

De ratio van de emotie

Gewogen loting is een rationeel systeem dat emoties oproept. Men vindt het oneerlijk. De vraag is echter of decentrale selectie eerlijker is. 'Decentrale selectie stelt instellingen in staat om de meest geschikte en gemotiveerde studenten voor de opleiding te selecteren en aspirant studenten hebben hun selectie voor de opleiding grotendeels zelf in de hand' is wat het kabinet vindt. Nu uiteenlopende selectieprocedures in de medische faculteiten worden ingevoerd is er echter minder zicht op hoe selectie wordt uitgevoerd. Het systeem wordt minder transparant. Er zit wat in de uitspraak 'de loting wordt nu vervangen door een casino'.⁷ Er zijn verschillende kansspelen, kandidaten weten niet bij welk spel hun kansen optimaal zijn, maar kunnen wel kiezen waar ze aan mee willen doen. Men neemt het lot wel meer in eigen hand, maar het is de vraag of dat echt eerlijker is. Een kwalitatieve selectieprocedure vraagt actieve inzet van kandidaten. Als inzet leidt tot toelating kan de kandidaat extra tevreden zijn en het gevoel hebben de toelating persoonlijk te hebben veroorzaakt. Dat is misschien de kern van de emotie. Als dat zo is, is selectie geen psychometrisch vraagstuk meer. Het gaan dan niet om een onafhankelijke meting van geschiktheid, maar om een onderscheid te maken tussen kandidaten die bereid zijn zich grote inspanning te getroosten en die dat niet zijn. Er is reden aan te nemen dat alleen al die inspanning om door een selectieprocedure te komen - ongeacht waar die op gericht is en wat er precies gemeten wordt - effect heeft op het latere studiesucces. Selectie is dan, in onderzoekstermen, niet zozeer een meetinstrument maar veeleer een interventie. Psychologen noemen dat een Hawthorne-effect.

Evidentie voor het Hawthorne effect bij selectie

Het Hawthorne-effect refereert aan een gedragsverandering die het gevolg is van het feit dat deelnemers aan een onderzoek weten dat zij worden geobserveerd.⁸ In het voorbeeld van figuur 1 is te zien dat studenten het beter deden als zij aan een selectieprocedure hadden meegedaan dan wanneer ze alleen via loting waren binnengekomen, *ongeacht* of ze waren geselecteerd of niet. Het is een aanwijzing dat deelname aan selectie zelf belangrijker is dan het selectie criterium. Een andere aanwijzing komt uit het promotieonderzoek van de Deense Lotte O'Neill. In Denemarken worden medische studenten toegelaten in twee groepen (Quota 1 en Quota 2): de helft van de plaatsen is gereserveerd voor de studenten met de hoogste cijfers op het vwo (Quota 1), de andere helft voor studenten die door een selectieprocedure zijn gegaan (een kennistest, de biografie, een motivatiebrief en een selectiegesprek). O'Neill keek naar *drop out* uit de medische opleiding tussen 2002 en 2007 en vond dat de kans daarop significant en aanzienlijk

(odds-ratio 0,58-0,61) kleiner was bij Quota 2 studenten. Binnen de Quota 2 groep bleek geen van de vier selectiecriteria voorspellend voor *drop out*. Zij concludeert: '*Selection by admission-testing appeared to have an independent, protective effect on drop out in this setting despite the lack of association between admission test scores and dropout.*'⁹

Het lijkt erop dat de waarde van selectie minder ligt in de betrouwbaarheid en validiteit en dat primair het feit dat kandidaten deelnemen een gunstig effect heeft op latere studieresultaten. Selectieprocedures kunnen motivatie niet goed *meten*, maar misschien wel *veroorzaken*. Bij meting *nadat* studenten waren toegelaten vonden Hulsman et al. dat geselecteerden hogere motivatiescores vertonen en meer inzet, maar niet betere studieprestaties dan studenten die via loting of de 8+ regeling toegelaten waren.¹⁰

Selectie is niet zozeer een meetinstrument, maar een interventie

Zware selectieprocedures voor medische opleidingen die veel vragen van de kandidaten, en die voor al in het buitenland veel voorkomen, *vormen* waarschijnlijk veel eerder dan dat ze bestaande eigenschappen van kandidaten *meten*. De opleidingen, zowel als de toegelaten kandidaten, kunnen zich na afloop tevreden voelen. Er wordt een beroep gedaan op gevoelens als 'ik pas blijkbaar goed bij deze opleiding', 'ze willen mij graag hebben', 'ik voel me nu loyaal aan deze instelling en gestimuleerd om hard aan de slag te gaan' en 'we hebben de beste studenten in huis gehaald, kijk hoe gemotiveerd ze blijken te zijn'.

De gedachte dat dat gebaseerd is geweest op een onafhankelijke, valide meting kan echter niet goed door evidentie gestaafd worden. Is dat erg? Misschien niet. De positieve gevoelens na afloop zijn gunstig voor het onderwijsklimaat. Wat echter niet terecht is, is dat afgevoerde kandidaten negatieve gevoelens behouden zoals 'ik ben blijkbaar niet geschikt', 'we zijn blij dat we de ongeschikte kandidaten hebben kunnen afwijzen'. Kandidaten die in een selectieprocedure terechtkomen gaan zich anders gedragen, en het blijft de vraag of dat nu de 'besten' zijn.

Conclusie

Selectieprocedures meten eigenschappen van kandidaten niet goed, en waar er wel voorspellende waarde wordt gevonden, neemt die af in de loop van de studie tot vrijwel betekenisloze verbanden. Dat is ook niet onlogisch. Onderwijs heeft immers tot taak om studenten tot nieuwe kennis, nieuwe vaardigheden en nieuw gedrag te brengen. Jonge mensen moeten zich kunnen ontplooiën en veranderen, kattenkwaad afleren, motivaties voor sommige zaken gaan ontwikkelen en voor andere zaken misschien verliezen. Het is de vraag of het gewenst is deze toekomst te willen voorspellen op achttienjarige leeftijd.

Deelname aan een selectie op zich heeft een gunstig effect op studieresultaten

Selectie vormt misschien eerder dan het meet. Voor diegenen die de inzet van een zware selectieprocedure willen leveren wil de faculteit zich waarschijnlijk graag inzetten, om ze te helpen door de studie te komen. Het waarnemen van die inzet kan in een korte selectieperiode voorafgaand aan de opleiding maar misschien ook in het begin van de opleiding zelf. In 1972, het jaar dat het lotingssysteem werd ingevoerd, stelde gezaghebbend psycholoog A.D. de Groot voor de propedeuse selectief te maken en de studenten die daar doorheen komen intensief naar de eindstreep te geleiden in wat hij noemde een expeditiemodel, in contrast met het bestaande veldloopmodel waarmee hij een geleidelijke afvalrace gedurende opleiding bedoelde.¹¹ Observeren van beginnende studenten gedurende een heel jaar is wel erg veel langer dan gedurende een kort selectieproces, maar enkele maanden studenten meemaken in de studie en dan de grootste inzet belonen met definitieve toelating is een model waarover kan worden nagedacht. Dat model zou er als volgt uit kunnen zien: Een grote groep kandidaten wordt bij de poort via een decentrale lotingsprocedure gereduceerd tot een groep die iets groter is, bijvoorbeeld 10 procent, dan wat uiteindelijk kan worden toegelaten. Voorselectie via motivatiebrieven of cv's kan ook, maar heeft weinig meerwaarde, is minder transparant en kost meer inspan-

ning. De groep die aan de studie begint, wordt na enkele maanden definitief geselecteerd op basis van getoonde inzet voor de studie. Deze selectie vormt de student, in activiteiten die al tot de studie behoren, en geeft de faculteit rustig de kans waar te nemen hoe de studenten passen bij de studie. Er is voor studenten genoeg prikkel om de inzet te tonen die de faculteit graag ziet en slechts voor 10 procent van hen een afwijzing. Dit is een hypothetische selectievariant waarover nagedacht kan worden, maar die redelijk aansluit bij de evidentie waarover wij beschikken.

Op 17 februari 2012 besloot het Kabinet het systeem van gewogen loting voor toelating tot opleidingen met een numerus fixus af te schaffen en gaf daarmee uitvoering aan een voornemen uit het regeerakkoord. Hiermee komt er een eind aan een tijdperk. Wat aanvankelijk werd gezien als een innovatief en zeer acceptabel systeem, wordt nu beschouwd als ondeugdelijk en oneerlijk. Er is een nieuwe politieke en maatschappelijke werkelijkheid die ons dwingt verder over selectie na te denken.

Prof dr Th.J. ten Cate

is werkzaam bij het expertisecentrum voor Onderwijs en Opleidingen UMC Utrecht

Noten:

- Schippers EI, Zijlstra H. Brief aan de Tweede Kamer betreffende Capaciteit en selectie opleidingen in de gezondheidszorg. 17 februari 2012.
- Gebaseerd op de voordracht Selection of medical students – can we do better than a lottery? Perspective from the Netherlands, gehouden voor de Association for the Study of Medical Education op 22 september 2011 te York, UK.
- Drenth PJD. In Nederland is selectie onmogelijk. Duykerlezing NRC Handelsblad 30 maart 1995.
- Salvatori P. Reliability and Validity of Admissions Tools Used to Select Students for the Health Professions. *Advances in Health Sciences Education* 2001;(6):159-175
- Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education* 2009; 43: 767-775
- Cohen-Schotanus J. Tegenintuïfief. Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar onderzoek van onderwijs in de medische wetenschappen aan de Rijksuniversiteit Groningen op dinsdag 2 november 2010
- Deze uitspraak is van Prof Marijke van Dijk, opleidingsdirecteur van de opleiding geneeskunde van de Universiteit Utrecht.
- Ball S. Untended Effects in Educational Research. In Keeves JP (ed). *Educational Research, Methodology and Measurement: an International Handbook*, 2e druk. Pergamon/Elsevier, Oxford, 1997, p 244.
- O'Neill L. Admission Tests in Medical Education. Generalizability, Predictive Validity and Dropout. PhD thesis University of Southern Denmark, 2011
- Hulsman RL et al. Effectiveness of selection in medical school admissions: evaluation of outcomes among freshmen. *Medical Education* 2007;41:369-377
- De Groot AD. Selectie voor en in het hoger onderwijs. Een probleemanalyse. 's-Gravenhage: Staatsuitgeverij, 1972