

Predictieve validiteit en sekse-specificiteit van het toelatingsexamen 'Arts en Tandarts' in Vlaanderen: Een eerste peiling.

F. Lievens, P. Coetsier, P.J. Janssen en C. Decaestecker

Samenvatting

In 1997 is in Vlaanderen het systeem van open toelating tot de medische faculteiten verlaten. In dit artikel rapporteren wij de resultaten van een eerste peiling naar de predictieve validiteit en de sekse-specificiteit van het ingevoerde toelatingsexamen. De steekproef in deze studie bestaat uit 941 kandidaten (359 mannen, 582 vrouwen; gemiddelde leeftijd = 18 jaar en 3 maanden), die deelnamen aan vier specifieke proeven en een casusgedeelte, dat ook bestond uit vier proeven. Als criteriumgegevens fungeren de eindresultaten van de eerste kandidatuur geneeskunde en tandheelkunde van het academiejaar 1997-1998 aan alle Vlaamse universiteiten. De correlatie tussen de totaalscore op het toelatingsexamen en deze eindscore bedraagt .35. De 'Redeneerproef' is de beste voorspeller. Hoewel zich op sommige proeven geslachtsgebonden verschillen in de gemiddelde scores voordoen, verschillen mannen en vrouwen niet significant van elkaar op de totaalscore op dit toelatingsexamen. Er is ook geen sprake van differentiële predictie.

Inleiding

Vanaf het academiejaar 1997-1998 dienen studenten die in Vlaanderen de universitaire opleiding geneeskunde en tandheelkunde willen gaan studeren te slagen in een toelatingsexamen. Op deze manier wil de Vlaamse Gemeenschap het aantal kandidaten, dat de opleiding aanvangt, beheersen (Ministerie van de Vlaamse Gemeenschap, 1996), zodat wie tot de opleiding toegang krijgt later daadwerkelijk het beroep van arts of tandarts zal kunnen uitoefenen.

Het Vlaamse toelatingsexamen 'Arts en Tandarts' bestaat uit twee delen: (1) kennis en

inzicht in de wetenschappen (KIW) en (2) informatieverwerven en -verwerken (IVV). Het eerste deel toetst de kennis en het inzicht in fysica, scheikunde, wiskunde en biologie. Dit deel werd na een arrest van het Belgische Arbitragehof van 27 mei 1997 geschorst voor het academiejaar 1997-1998. De reden hiervoor was dat het toelatingsexamen nog niet bestond op het moment waarop deze studenten als leerling aan de laatste cyclus van hun secundair onderwijs begonnen. Daarom hadden zij hun keuze van afstudeerrichting niet kunnen afstemmen op (dit deel van) het toelatingsexamen. Het deel IVV bleef behouden. Enerzijds bestaat dit uit vier specifieke proeven, met name een 'Redeneerproef', een proef 'Inzicht in schema's en figuren', een geheugenproef 'Feiten leren' en een proef 'Patroonherkenning' (Minnaert, 1996; Stinissen & Vander Steene, 1980). Anderzijds omvat dit deel ook vier ad hoc geconstrueerde proeven, namelijk een op video opgenomen 'Les', een 'Stilleestekst' ('silent reading protocol'), een op video opgenomen 'Arts-patiënt gesprek' en een op video opgenomen 'Teambespreking' (Coetsier & Lievens, 1997; Lievens & Coetsier, 1998). Over al deze onderdelen moeten nadien telkens aansluitend meerkeuzevragen worden beantwoord.

Het arrest van het Arbitragehof illustreert de maatschappelijke beroering, die het instellen van dit toelatingsexamen 'Arts en Tandarts' in Vlaanderen heeft teweeggebracht. Tot op heden ontbreekt - het kan moeilijk anders - een wetenschappelijke evaluatie van dit toelatingsexamen. Veel vragen blijven dan ook onbeantwoord. Wat is de voorspellende kracht van dit toelatingsexamen? Zijn er verschillen tussen mannen en vrouwen op de diverse onderdelen van het examen? Hoe bekijken de kandidaten het examen? Is het mogelijk om er zich op voor te bereiden?

In dit artikel rapporteren wij de resultaten van een eerste peiling naar de predictieve validiteit en de sekse-specificiteit van dit toelatingsexamen. Drie onderzoeksvragen staan centraal:

Als eerste onderzoeksvraag gaan wij na in welke mate het toelatingsexamen een goede voorspeller vormt voor het eindresultaat in 'de' eerste kandidatuur geneeskunde en tandheelkunde aan alle Vlaamse universiteiten.

Als tweede onderzoeksvraag analyseren wij welke onderdelen van het toelatingsexamen als beste voorspellers naar voren treden.

Als derde onderzoeksvraag gaan wij na of mannen en vrouwen gemiddeld verschillend scoren op het toelatingsexamen en of deze eventuele verschillen ook gerelateerd zijn aan gelijkaardige verschillen in het eindresultaat in de eerste kandidatuur geneeskunde en tandheelkunde.

1. Het toelatingsexamen 'Arts en Tandarts' in Vlaanderen

Vele internationale studies laten zien dat cognitieve voorspellers zoals voorgaande studieresultaten (Tomlinson, Clack, Pettingale, Anderson & Ryan, 1977; Green, Peters & Webster, 1993; Smal, 1995), kennis van wetenschapsvakken (McManus & Richards, 1986; Montague & Odds, 1990) en (non-) verbale en numerieke intelligentie (Roessler, Lester, Butler, Rankin & Collins, 1978) goede voorspellers zijn voor studiesucces in het medisch onderwijs. Omdat er in Vlaanderen geen nationaal examen bestaat op het einde van het secundair onderwijs ontbreken voorgaande studieresultaten als voorspeller. Daarom was een eerste aandachtspunt bij het ontwerp van het Vlaams toelatingsexamen 'Arts en Tandarts' (versie 1997) dat specifieke proeven, die peilden naar de intelligentie en de cognitieve vaardigheden van de studenten, een belangrijk onderdeel dienden te vormen van dit examen. Er werd gekozen voor proeven waarvan betrouwbaarheids- en validiteitsgegevens beschikbaar waren op basis van vroeger onderzoek bij generatiestudenten in de geneeskunde (Minnaert, 1996) en in andere universitaire eerstejaarsopleidingen in Vlaanderen (Stinissen & Vander Steene,

1980). De eerste specifieke proef, de 'Redeneerproef', bestond uit meerkeuzevragen met een in moeilijkheidsgraad opklimmende serie van verbale, numerieke en diagrammatische opdrachten. Wegens de goede voorspellende kracht (Stinissen & Vander Steene, 1980; zie ook: Minnaert, 1996) woog deze proef meer door in de totaalscore op het toelatingsexamen (zie Tabel 1). In de tweede proef 'Inzicht in schema's en figuren' werd de visuo-spatiale scanningsmogelijkheid getest. De kandidaat diende immers snel en accuraat schema's te lezen en te interpreteren. In de geheugenproef 'Feiten leren' stond de vaardigheid om informatie correct te onthouden en na zekere tijd te reproduceren centraal. De vierde en laatste specifieke proef 'Patroonherkenning' verwachtte van de studenten dat ze voor elk complex patroon bepaalden welke van vijf enkelvoudige stimulusfiguren er deel van uitmaakte.

Naast het gedeelte specifieke proeven werd ook in functie van dit toelatingsexamen 'Arts en Tandarts' een casusgedeelte ontwikkeld. Lievens en Coetsier (1998) beschrijven uitvoerig de verantwoording, de opbouw en de validatie via expert artsen van dit casusgedeelte. In het eerste onderdeel van dit casusgedeelte werd een op video opgenomen hoorcollege over het koolhydraatmetabolisme getoond. Tijdens deze 'Les' konden de kandidaten, zoals studenten genees- en tandheelkunde, aantekeningen maken om nadien de vragen te kunnen oplossen. De tweede proef was een 'Stilleestekst'. De studenten konden bij het beantwoorden van de hierbij horende vragen zo dikwijls als nodig de tekst raadplegen. In deze twee 'miniaturproeven' werden de kandidaten geconfronteerd met situaties die representatief en relevant waren voor universitair onderwijs (i.e. het volgen van een hoorcollege en het volledig zelf grondig instuderen van een tekst). De verwachting bij deze proeven was dan ook dat de wijze waarop de kandidaat hiermee omgaat voorspellend zou zijn voor zijn/haar toekomstig gedrag in het universitair medisch onderwijs.

Het is duidelijk dat de vier specifieke proeven en deze twee proeven van het casusgedeelte eerder cognitief georiënteerd waren. Ondanks het veelvuldig gebruik en het be-

lang van deze cognitieve factoren tonen zij slechts één zijde van de medaille (Ten Cate, Van Rossum & Van der Vleuten, 1997). Cognitieve factoren omvatten immers slechts 35% van de te verklaren variantie in de academische prestaties van medische studenten (Walton, 1987). Een andere beperking is dat cognitieve factoren in hogere mate samenhangen met de prestaties van medische studenten in preklinische jaren (Baneke, 1989; Green, Peters & Webster, 1991; Smal, 1995). Daarom bevatte het casusgedeelte van het toelatingsexamen nog twee andere proeven die een weerspiegeling vormden van twee karakteristieke situaties uit het werkdomein van artsen: het 'Arts-patiënt gesprek' en de 'Teambespreking'. Deze op video opgenomen proeven peilden hoe de kandidaten op deze meer klinische situaties reageerden.

Aangezien de vier proeven van het casusgedeelte speciaal ontwikkeld werden voor dit toelatingsexamen ontbraken vooralsnog validiteitsgegevens bij studenten geneeskunde. Zoals reeds gezegd was dit niet het geval bij de specifieke proeven. Daarom fungeerden deze proeven volgens de ontwerpers van het toelatingsexamen als 'ankerpunten' om de validiteit van het casusgedeelte af te leiden.

2. Methode

2.1 Steekproef

In juli 1997 namen 941 kandidaten (359 mannen, 582 vrouwen; respectievelijk 38.2% en 61.8%) deel aan de eerste sessie (juli 1997) van het toelatingsexamen 'Arts en Tandarts'. De gemiddelde leeftijd van de deelnemers bedroeg 18 jaar en 3 maanden.

Voor onze studie naar de predictieve validiteit van het toelatingsexamen konden wij enkel beroep doen op personen die tijdens het aansluitend academiejaar (1997-1998) daadwerkelijk gestart waren in de eerste kandidatuur geneeskunde en tandheelkunde en die er een officieel eindresultaat behaalden. Dit waren 623 kandidaten (235 mannen, 388 vrouwen). Van hen volgden 567 personen de richting geneeskunde en 56 de richting tandheelkunde. 558 studenten hadden de Belgische, 65 de Nederlandse nationaliteit.

2.2 Procedure van het toelatingsexamen

Het toelatingsexamen 'Arts en Tandarts' werd tijdens twee opeenvolgende voormiddagen plenair georganiseerd in een grote zaal te Brussel. De eerste voormiddag stonden de vier specifieke proeven op het programma. Na een korte verwelkoming en de algemene instructies kregen de kandidaten als eerste opdracht ('Feiten leren') een geheugenkaart met instructieblad voorgelegd. Op de geheugenkaart werden 15 patiënten voorgesteld (in 5 groepen van 3) met vermelding van hun naam, leeftijd, beroep, een beschrijvend kenmerk en hun aanmeldingsklacht of diagnose. De kandidaten kregen 6 minuten om deze 15 personen in te prenten. Daarna werd de geheugenkaart opgehaald. Vervolgens werd de 'Redeneerproef' uitgedeeld. De kandidaten hadden 45 minuten tijd om de 54 items in deze proef af te werken. De volgende proef was 'Inzicht in schema's en figuren'. De kandidaten kregen 14 minuten werktijd om de 32 vragen van deze proef op te lossen. Nadien werd gepeild naar de reproductie van de ingeprente patiënten-beschrijvingen. Er was 7 minuten tijd om 20 vragen daarover te beantwoorden. De eerste dag werd afgesloten met de proef 'Patroonherkenning' (52 vragen, 30 minuten). Zoals in de voorgaande specifieke proeven waren ook dit meerkeuzevragen met vijf antwoordalternatieven.

De tweede voormiddag stond het casusgedeelte centraal. Dit werd afgewerkt met behulp van een werkboek, dat de instructies, contextuele informatie en vragen bevatte. De eerste opdracht bestond uit het volgen van een op video opgenomen 'Les' (duur = 25 minuten). Zoals in een hoorcollege gebruikelijk is, mochten de kandidaten aantekeningen maken en kregen zij de door de docent gebruikte kopieën van de transparanten. Nadien hadden zij 25 minuten de tijd om met behulp van hun aantekeningen 40 vragen op te lossen. Daarna werd een 'Stillestekst' voorgelegd over een medisch onderwerp; daarbij werden 20 vragen ter oplossing aangeboden (werktijd = 40 minuten). De kandidaten kregen de aanbeveling de tekst eerst volledig te lezen en pas nadien de vragen te beantwoorden. Daarbij konden zij, zo dikwijls als nodig, de tekst raadplegen. De derde opdracht bestond uit een op video opgenomen

'Arts-patiënt gesprek'. Allereerst verschenen pasfoto's van de patiënt en de arts in beeld. Aansluitend werd de patiëntenfiche getoond en kort toegelicht. Daarna werd de reden van het bezoek van de patiënt aan de arts geschetst. Vervolgens bekeken de kandidaten het gesprek (duur = 6 minuten). Hierbij mochten ze zoveel aantekeningen maken als ze wilden. Daarna kregen ze 25 minuten tijd om hierover 30 vragen op te lossen. Tot slot van het casusgedeelte kregen kandidaten een 'Teambespreking' over een (andere) patiënt te zien. Vooraf werden de patiëntenfiche, de reden van opname, de sociale achtergrond, de huidige situatie en de teamleden voorgesteld. Daarna bekeken de kandidaten het videofragment (duur = 11 minuten). Vervolgens kregen ze 25 minuten tijd om hierover 30 vragen op te lossen. Alle vragen bij dit casusgedeelte waren van het meerkeuzetype met twee tot vier antwoordalternatieven.

2.3 Meetinstrumenten

Voorspellers

Voor elk van de acht proeven van het toelatingsexamen werd een totaalscore berekend door het aantal juiste antwoorden op te tellen. Er werd telkens een milde gokcorrectie toegepast. Vervolgens werden een gewogen som van de vier specifieke proeven en een gewogen som van de vier proeven van de casus berekend (zie Tabel 1 voor de specifieke gewichten). Deze gewichten waren vooraf bij wetgeving bepaald. De som van deze twee gewogen scores resulteerde dan in de totaalscore op het toelatingsexamen. Om te slagen voor het toelatingsexamen dienden de kandidaten ten minste 12 op 20 te behalen.

Criterium

In dit onderzoek fungeerden de eindresultaten van de eerste kandidatuur geneeskunde en tandheelkunde als criterium. Specifiek betrof het de eindresultaten van de eerste examenperiode (juli 1998) van de studenten aan de vijf Vlaamse universiteiten uit het academiejaar 1997-1998. Studenten konden maximaal een score van 20 behalen.

Een mogelijk probleem bij dit criterium was dat in elke universiteit deze eindscore op een verschillende manier werd bepaald omdat de feitelijke invulling van deze eerste kandi-

datuur (wat betreft vakken en professoren) verschilde van universiteit tot universiteit. Daarom voerden wij de analyses zowel over alle universiteiten heen als per universiteit uit.

3. Resultaten

3.1 Beschrijvende statistieken

Tabel 1 presenteert de beschrijvende statistieken van de specifieke proeven, de casus en de totaalscore op het toelatingsexamen. Verhoudingsgewijs lagen de scores op de specifieke proeven gemiddeld veel lager dan de respectievelijke scores op de casus. De standaarddeviaties van de specifieke proeven waren groter dan de standaarddeviaties van de afzonderlijke casusonderdelen.

Voor het toelatingsexamen slaagden 688 (72.58%) van de 941 kandidaten.

De interne consistentiecoëfficiënten van de specifieke proeven waren aanvaardbaar (variërend van .70 tot .92). Dit was niet het geval voor de proeven van de casus (van .41 tot .56). Het is echter zo dat voor situationele tests interne consistentie (in tegenstelling tot test-hertest betrouwbaarheid) niet echt een gepaste betrouwbaarheidsmaat is (Motowidlo & Tippins, 1993).

3.2 Predictieve validiteit

In het kader van de eerste onderzoeksvraag berekenden wij Pearson-correlaties tussen de scores op de verschillende onderdelen van het toelatingsexamen en het eindresultaat in de eerste kandidatuur. Deze correlaties bevinden zich in Tabel 2 en werden gecorrigeerd voor 'direct restriction of range' (Thorndike, 1949).

Tabel 2 toont dat de totaalscore op het toelatingsexamen .35 correleerde met het geregistreerde eindresultaat in de eerste kandidatuur (over alle universiteiten heen, zie eerste kolom van Tabel 2). Om de waarde van deze correlatie beter te kunnen inschatten, is het interessant om te kijken hoeveel van de 'toegelaten' studenten ook daadwerkelijk geslaagd waren in de eerste kandidatuur. Na de eerste examenperiode bedroeg dit percentage 56.76%. Na de tweede examenperiode steeg het tot 72.95%. Wanneer wij deze correlatie berekenen per universiteit, was er nogal wat variatie in de resul-

Tabel 1

Beschrijvende statistieken van de variabelen van deze studie.

	\bar{X}	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Geslacht	--	--	--											
2. Redeneerproef (54 ^a , .50 ^b)	26.27	5.94	-.05	-. ^c										
3. Inzicht in schema's (32, .20)	11.76	5.22	-.13**	.28**	.83									
4. Geheugenproef (20, .10)	9.10	3.86	.14**	.24**	.03	.70								
5. Patroonherkenning (52, .20)	23.86	10.80	-.05	.31**	.29**	.09**	.92							
6. Les (40, .33)	30.26	4.31	-.01	.31**	.14**	.15**	.17**	.55						
7. Stilleestekst (20, .17)	15.18	2.37	.05	.33**	.10**	.16**	.13**	.37**	.56					
8. Gesprek arts-patiënt (30, .25)	23.45	3.01	.16**	.25**	.02	.19**	.07*	.29**	.31**	.41				
9. Teambespreking (30, .25)	24.17	2.92	.08*	.22**	.06	.21**	.05	.29**	.22**	.32**	.48			
10. Totaal specifieke proeven (10) ^d	4.54	1.00	-.07*	.81**	.61**	.37**	.69**	.31**	.30**	.21**	.20**	--		
11. Totaal casus (10) ^d	7.76	.73	.09**	.40**	.12**	.25**	.16**	.79**	.63**	.67**	.64**	.37**	--	
12. Totaal toelatingsexamen (20)	12.30	1.45	.00	.77**	.48**	.39**	.56**	.62**	.53**	.49**	.46**	.88**	.76**	--
13. Eindscore eerste jaar (20)	11.14	3.48	.02	.30**	-.03	.13**	.11**	.19**	.19**	.10**	.02	.24**	.20**	.29**
				(.33)	(-.03)	(.12)	(.12)	(.20)	(.21)	(.12)	(.02)	(.27)	(.23)	(.35)

Noot. N varieert tussen 609 en 941. Interne consistentiecoëfficiënten (alpha's) bevinden zich op de diagonaal. Correlaties tussen haakjes werden gecorrigeerd voor 'direct restriction of range'. Statistische significantie werd bepaald vooraleer de correctie voor restriction of range werd toegepast.

^a Dit cijfer verwijst naar het aantal items (of de maximale score) per proef.

^b Dit cijfer verwijst naar het gewicht van deze proef bij de berekening van de gewogen totaalscores (van de specifieke proeven en van het casusgedeelte).

^c Door de scherpe tijdslimiet waren er te weinig personen die de proef volledig invulden. Hierdoor kon de alpha niet berekend worden.

^d Deze scores werden verkregen door de scores op de afzonderlijke proeven, die elk eerst vermenigvuldigd werden met het betreffende gewicht, bij elkaar op te tellen.

* $p < .05$; ** $p < .01$.

taten. Zo was de predictieve validiteit van de totaalscore op het toelatingsexamen in één universiteit .22 en in een andere universiteit .56.

Uit Tabel 2 blijkt ook een significante correlatie (gemiddeld .27 met schommelingen van .17 tot .49 per universiteit) tussen de specifieke proeven en het eindresultaat eerste kandidatuur. De correlatie van de casus met het eindresultaat was altijd lager en schommelde tussen de .11 en .43. Deze correlatie was ook slechts in twee universiteiten significant. Tabel 2 toont ten slotte dat van alle afzonderlijke proeven van het toelatingsexamen de 'Redeneerproef' het best scoort. Over alle universiteiten heen bedraagt de correlatie .33 met het eindresultaat eerste kandidatuur. Per universiteit schommelt de correlatie tussen .24 en .57. Enkele uitzonderingen uitgesloten (bijvoorbeeld geheugenproef, les en stilleestekst) zijn de correlaties van de ande-

re proeven met het eindresultaat in de eerste kandidatuur aan de lage kant.

In het kader van onderzoeksvraag 2 gingen wij via hiërarchische regressie na welke proeven van het toelatingsexamen naar voren traden als beste voorspellers. De vier specifieke proeven werden eerst in de regressievergelijking gebracht, gevolgd door de vier onderdelen van het casusgedeelte. Tabel 3 toont de resultaten. Als we de analyses uitvoeren over alle universiteiten heen, verklaarden de vier specifieke proeven ongeveer 10 percent van de variantie. Hierbij trad de 'Redeneerproef' ($\beta = .26, p < .01$) naar voren als belangrijkste voorspeller. Het regressiegewicht voor de proef 'Inzicht in schema's en figuren' was statistisch significant maar negatief ($\beta = -.09, p < .05$). De vier onderdelen van het casusgedeelte droegen significant (i.e. een extra 4 percent) bij tot verhoging van

Tabel 2

Correlaties tussen proeven toelatingsexamen en eindscore: algemeen beeld en uitgesplitst per universiteit (N varieert^a van 59 tot 201).

	Algemeen	Uitsplitsing per universiteit				
		1	2	3	4	5
Redeneerproef	.30** (.33)	.27** (.30)	.48** (.54)	.37** (.40)	.50** (.57)	.23** (.24)
Inzicht in schema's	-.03 (-.03)	-.05 (-.05)	.05 (.05)	-.01 (-.02)	.00 (.00)	-.03 (-.04)
Geheugenproef	.13** (.12)	.13 (.12)	.25 (.27)	.09 (.09)	.30** (.29)	.08 (.09)
Patroonherkenning	.11** (.12)	.18* (.18)	.18 (.19)	.16 (.18)	.15 (.16)	.07 (.07)
Les	.19** (.20)	.28** (.30)	.23 (.26)	.06 (.05)	.12 (.14)	.13 (.14)
Stilleestekst	.19** (.21)	.22** (.27)	.21 (.24)	.22 (.18)	.11 (.14)	.15* (.16)
Gesprek arts-patiënt	.10** (.12)	.07 (.07)	.21 (.25)	.18 (.23)	.08 (.09)	.08 (.09)
Teambespreking	.02 (.02)	.08 (.08)	.23 (.24)	.05 (.06)	.06 (.07)	-.15* (-.17)
Totaal specifieke proeven	.24** (.27)	.25** (.29)	.40** (.44)	.30* (.34)	.43** (.49)	.16* (.17)
Totaal casus	.20** (.23)	.25** (.28)	.36* (.43)	.17 (.17)	.14 (.19)	.09 (.11)
Totaal toelatingsexamen	.29** (.35)	.32** (.38)	.48** (.56)	.34** (.41)	.41** (.52)	.17* (.22)

Noot. Correlaties tussen haakjes werden gecorrigeerd voor restriction of range. Statistische significantie werd bepaald vooraleer de correctie voor restriction of range werd toegepast.

^a Om de anonimiteit van de universiteiten enigszins te waarborgen vermeldden wij niet de exacte N waarop de correlaties berekend werden.

* $p < .05$; ** $p < .01$.

de verklaarde variantie. Dit betekent dat het toelatingsexamen in zijn geheel ten opzichte van het criterium 14% van de variantie verklaarde. De 'Stilleestekst' ($\beta = .11, p < .01$) en de 'Les' ($\beta = .11, p < .01$) fungeerden als significante voorspellers in het casusgedeelte.

Wanneer we de resultaten uitsplitsen per universiteit verandert het beeld. Zo is de redeneerproef de enige proef die de eindresultaten eerste kandidatuur in elke universiteit voorspelt. Andere proeven zoals de les en de stilleestekst zijn voorspellend voor de eindresultaten in één universiteit maar niet voor de eindresultaten van een andere universiteit. Hierdoor droegen de casusproeven slechts in drie universiteiten bij tot een significante verhoging van de verklaarde variantie (zie onderste lijnen van Tabel 3).

Aangezien het probleem van range restrictie ook kan meespelen in regressie-analyses, pasten we de multivariate range restrictie correctie van Lawley (1943, geciteerd in Bobko, 1995; Ree, Carretta, Earles & Albert, 1994) toe op de volledige matrix van correlaties. Deze 'gecorrigeerde' matrix fungeerde dan telkens als input voor hiërarchische regressie-analyses. De 'gecorrigeerde' R^2 's waren slechts lichtjes hoger dan in de zopas gerapporteerde analyses. Bovendien waren de regressiegewichten gelijklopend met bovenstaande resultaten.

3.3 Sekse-specificiteit

De derde onderzoeksvraag handelde over mogelijke sekse-verschillen in de gemiddelde totaalscore op het toelatingsexamen en op

Tabel 3

Regressiegewichten van proeven van toelatingsexamen op totaalscore eerste kandidatuur geneeskunde en tandheelkunde: algemeen beeld en uitgesplitst per universiteit (N varieert van 59 tot 201).

Proef	Algemeen	1	2	3	4	5
Stap 1						
Redeneerproef	.29**	.24**	.45**	.36**	.47**	.24**
Inzicht in schema's	-.09*	-.09	-.04	-.09	-.10	-.10
Geheugenproef	.06	.08	.16	.02	.13	.03
Patroonherkenning	.06	.14	.03	.08	.14	.03
Stap 2						
Redeneerproef	.26**	.17*	.44**	.46**	.47**	.18*
Inzicht in schema's	-.09*	-.09	-.03	-.08	-.10	-.10
Geheugenproef	.05	.05	.13	-.06	.13	.02
Patroonherkenning	.06	.13	.01	.11	.12	.01
Les	.11**	.21**	.00	-.24	.05	.12
Stilleestekst	.11**	.12	.11	.42**	.07	.11
Gesprek arts-patiënt	.03	-.05	.21	.12	.00	.06
Teambespreking	-.06	-.03	.06	.08	-.05	-.20*
R^2 Stap 1	.10**	.10**	.26**	.15*	.30**	.06*
ΔR^2 voor Stap 2	.04**	.06**	.07	.14*	.01	.05*

* $p < .05$; ** $p < .01$.

de respectieve onderdelen daarvan. Tabel 4 splitst de resultaten op de verschillende proeven van het toelatingsexamen uit voor mannen en vrouwen.

Uit t -toetsen (zie de laatste kolom van Tabel 4) komt naar voren dat vrouwen significant hoger scoorden op het 'Arts-patiënt gesprek', de 'Teambespreking' en op het casusgedeelte in het algemeen. Ook op de geheugenproef 'Feiten leren' presteerden zij significant beter dan de mannen. Op 'Inzicht in schema's en figuren' scoorden deze laatsten dan weer significant hoger. Hoewel significante geslachtsgebonden verschillen werden gevonden, tonen de effectgroottes (zie laatste kolom van Tabel 4) dat deze verschillen niet erg groot waren. De sekse-specifieke verschillen neutraliseerden elkaar ook aangezien beide geslachten gemiddeld niet significant verschilden in de totaaluitslag op het toelatingsexamen. Dit bleek eveneens uit een vergelijking van de slaagpercentages van mannen en vrouwen (respectievelijk 71.59% en 72.85%).

Wij bestudeerden de sekse-specificiteit van het toelatingsexamen ook via een meer genuanceerde techniek. Wij onderzochten namelijk of er sprake was van differentiële predictie: Waren de gevonden verschillen tussen mannen en vrouwen in gemiddelden

gerelateerd aan verschillen in het eindresultaat in de eerste kandidatuur geneeskunde en tandheelkunde? Bij deze analyse volgden wij het Cleary-model (1968) inzake test bias, dat door psychometrici algemeen wordt aanvaard (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1998; Society for Industrial and Organizational Psychology, 1987). In het bijzonder voerden wij een hiërarchische regressie uit, waarbij als eerste blok alle acht proeven van het toelatingsexamen werden ingevoerd (zie Bartlett, Bobko, Mosier & Hannon, 1978). De eindscore in de eerste kandidatuur fungeerde als criterium. Logischerwijze waren deze resultaten dezelfde als de resultaten in Tabel 3. Als tweede blok voerden wij de variabele geslacht in en als derde en laatste blok de acht producttermen van elke proef met geslacht. Geslacht bleek geen significante voorspeller nadat de effecten van de voorspellers eruit genomen waren. Er was dus geen sprake van een zogenaamde 'intercept bias', hetgeen betekent dat er geen verschillen bestonden tussen beide geslachten op hun eindresultaat in de eerste kandidatuur, nadat zij gelijkgesteld waren op het gebied van de bekwaamheid, die gemeten werd door de proeven. Ook de producttermen van geslacht

Tabel 4

Gemiddelden (\bar{X}) en standaarddeviaties (SD) van de proeven van het toelatingsexamen voor mannen en vrouwen

Proef	Mannen ($N = 359$)		Vrouwen ($N = 582$)		t-waarde	Effect- grootte
	\bar{X}	SD	\bar{X}	SD		
Redeneerproef	26.62	6.34	26.06	5.68	1.36	-0.09
Inzicht in schema's	12.56	5.33	11.27	5.10	3.71**	-0.25
Geheugenproef	8.36	3.90	9.55	3.77	-4.64**	0.31
Patroonherkenning	24.40	10.76	23.53	10.82	1.20	-0.08
Les	30.27	4.46	30.25	4.22	.07	0.00
Stilleestekst	15.04	2.39	15.27	2.35	-1.44	0.10
Gesprek arts-patiënt	22.88	3.09	23.80	2.90	-4.53**	0.30
Teambespreking	23.89	2.97	24.34	2.88	-2.33**	0.16
Totaal specifieke proeven	4.61	1.06	4.50	0.97	1.58	-0.11
Totaal casus	7.67	0.77	7.81	0.71	-2.69**	0.18
Totaal toelatingsexamen	12.28	1.56	12.31	1.37	-.25	0.02

* $p < 0.05$; ** $p < 0.01$.

met de proeven waren statistisch niet significant. Er was dus ook geen sprake van een zogenaamde 'slope bias'. De proeven waren dus niet meer voorspellend voor de ene groep dan voor de andere. Wij voerden dergelijke hiërarchische regressie-analyses ook per universiteit uit. Er was nergens sprake van differentiële predictie.

4. Conclusies

De doelstelling van deze studie bestaat erin het Vlaams toelatingsexamen 'Arts en Tandarts' vanuit twee belangrijke graadmeters (i.e. predictieve validiteit en sekse-specificiteit) een eerste keer te evalueren. Op grond van deze studie kunnen wij de volgende conclusies trekken.

Allereerst ligt het aantal juiste antwoorden bij de specifieke proeven gemiddeld veel lager dan bij het casusgedeelte. Dit kan betekenen dat de specifieke proeven veel moeilijker zijn en/of dat de werktijd bij deze proeven veel beperkter is. Ondanks hun vrij lage scores op de specifieke proeven slagen vele studenten voor het toelatingsexamen dankzij hun goede scores op het casusgedeelte. Op die manier scoort de gemiddelde kandidaat uiteindelijk boven de bij wetgeving vooropgezette 'cut-off' van 12 op 20.

Het toelatingsexamen bezit verhoudingsgewijs een redelijke tot goede voorspellende kracht. Er bestaat tussen het volledige toelatingsexamen en 'de' totaaluitslag van de studenten eerste kandidatuur geneeskunde en

tandheelkunde meestal een substantiële correlatie van boven de .30. Concreet bedraagt het percentage 'toegelaten' studenten, dat nadien ook daadwerkelijk in de eerste kandidatuur was geslaagd, 56.76% (na de eerste examenperiode) en 72.95% (na de tweede examenperiode). Onze per universiteit uitgesplitste analyses laten echter ook gevoelige verschillen zien qua voorspellende validiteit. Zo behaalt het toelatingsexamen in één universiteit een voorspellende validiteit van .56 en in een andere universiteit slechts .22. De meest voor de hand liggende verklaring voor deze verschillen per universiteit is dat er in sommige universiteiten een grotere overeenkomst is tussen de examens (qua vorm, vragen en ondervraagde vakken) in het eerste jaar en de bekwaamheden gemeten in de tests en proeven van het toelatingsexamen. Zo verschillen Vlaamse universiteiten nogal naargelang de nadruk die er wordt gelegd op het belang van wetenschapsvakken in de eerste kandidatuur. Dit accentverschil kan ertoe leiden dat de redeneertest (en bij gevolg de totaalscore) een grotere voorspellende kracht heeft voor de eindresultaten in sommige universiteiten. Ook verschillen Vlaamse universiteiten qua onderwijsvorm. Zo is in sommige universiteiten het hoorcollege nog steeds de meest gebruikte onderwijsvorm. In andere (kleinere) universiteiten wordt geëxperimenteerd met een meer probleemgestuurde onderwijsvorm bij het doceren. In dit verband is het inderdaad opvallend dat de predictieve

validiteit hoger ligt bij kleinere universiteiten. In elk geval verdient het aanbeveling om in de toekomst voor alle kandidaten dezelfde maat als criterium te gebruiken. Dit kan bijvoorbeeld via een vooraf door een externe instantie geconstrueerde eindproef, die in alle universiteiten op een gestandaardiseerde wijze in de eerste kandidatuur afgenomen zou worden. Op dit moment is dit echter (nog) onmogelijk.

Behalve deze verschillen per universiteit zijn er nog twee andere elementen waardoor onze resultaten met enige omzichtigheid geïnterpreteerd dienen te worden. Allereerst betreft het hier een nog onvolledige versie van dit toelatingsexamen. Proeven over 'Kennis en inzicht in wetenschapsvakken', die algemeen een sterke voorspellende kracht hebben voor academische prestaties in de eerste jaren van het medisch onderwijs (McManus & Richards, 1986; Montague & Odds, 1990), waren voor het academiejaar 1997 immers afgevoerd van het examenprogramma. Ten tweede fungeerde in deze correlatiestudies enkel het eindresultaat in de eerste kandidatuur geneeskunde en tandheelkunde als criterium. Naast replicatie-onderzoek is dan ook longitudinaal onderzoek nodig om de stabiliteit van deze correlaties vast te stellen.

Binnen de specifieke proeven fungeert de 'Redeneerproef' als de beste voorspeller. Bovendien heeft de redeneertest in elke universiteit een significante correlatie met het eindresultaat in de eerste kandidatuur. Verschillende auteurs (Mitchell, Haynes & Koenig, 1994; Vu, Dawson-Sauders & Barrows, 1987; Minnaert, 1996) vonden gelijkaardige resultaten met betrekking tot redeneerproeven. De overige drie specifieke proeven daarentegen scoorden relatief zwak. Dit geldt in het bijzonder voor de proef 'Inzicht in schema's en figuren', die zelfs negatief correleert. Tegen de verwachtingen in hielden 'Feiten leren' (zie Patel, Groen & Frederiksen, 1986) en 'Patroonherkenning' (zie Wilson, Suddick, Shay & Hustmyer, 1981) ook weinig tot geen verband met de academische prestatie.

De 'Les' en 'Stilleestekst' blijken binnen het casusgedeelte voor sommige universiteiten goede voorspellers. Hun voorspellende

kracht ontlenen zij aan het feit dat de hoofdtaken van eerstejaarsstudenten bestaan uit het volgen van hoorcolleges en het instuderen van hiermee verband houdende teksten (Lip-ton, Huxham & Hamilton, 1984). Bovendien toont hiërarchische regressie dat voor sommige universiteiten beide niet volledig overlappen met de specifieke proeven; zij verklaren immers een extra deel van de variantie.

Het arts-patiënt gesprek en de teambespreking scoren voor nagenoeg alle universiteiten zwak. Beide registreren interpersoonlijke en communicatieve vaardigheden, waarin kandidaten zich nog niet hebben kunnen oefenen en profileren. Dergelijke vaardigheden komen immers pas aan bod tijdens de klinische jaren en in het beroep. Wij verwachten vooralsnog op termijn redelijke correlaties tussen gesprek en teambespreking enerzijds en de resultaten tijdens de klinische jaren anderzijds (zie Glaser, Hojat, Veloski, Blacklow & Goepf, 1992). Daarom is het nog voorbarig om dit arts-patiënt gesprek en deze teambespreking uit het toelatingsexamen te verwijderen.

Dit toelatingsexamen blijkt niet onderhevig aan differentiële predictie. Er is immers noch sprake van een 'intercept bias' noch van een 'slope bias'. Ook verschillen beide geslachten niet significant van elkaar qua totaalscore. Wel vonden wij sekse-specifieke verschillen in de gemiddelden van de afzonderlijke proeven. Deze komen echter overeen met bevindingen uit voorgaande studies (zie Jensen, 1998).

Wij sluiten af met mogelijke beleidsimplies van onze bevindingen. Indien men de instroom in de studies arts en tandarts wenst te beperken, dan toont deze eerste evaluatie dat een toelatingsexamen daartoe een goed instrument kan vormen. Zo is het bemoedigend dat, hoewel het onderdeel 'Kennis en inzicht in wetenschapsvakken' geen deel vormde van het toelatingsexamen, er toch reeds een belangrijk deel van de variantie wordt verklaard. Deze eerste correlaties vertonen verder een patroon dat overeenstemt met de verwachtingen. Dit neemt niet weg dat tegenstanders van dit toelatingsexamen bovenstaande resultaten kunnen gebruiken als 'het bewijs' voor hun stelling dat een dergelijk examen slechts 14% van de variantie in

studieprestaties in het eerste jaar verklaart. Een tweede implicatie is dat deze voorlopige resultaten tonen dat een toelatingsexamen 'Arts en Tandarts' in elk geval een redeneertest dient te omvatten. Ten derde is het belangrijk om de predictoren niet te beperken tot cognitieve proeven. Daarom bevatte het Vlaams toelatingsexamen het casusgedeelte, dat bestond uit op video opgenomen proeven. Hoewel deze videosimulaties een hoge kostprijs en (voorlopig) weinig predictieve kracht hebben, verdienen zij toch hun plaats in een toelatingsexamen. Voordelen ervan zijn namelijk dat de proeven beogen om naar andere vaardigheden te peilen dan cognitieve vaardigheden en dat kandidaten enthousiast reageren op deze proeven. Bovendien bevestigt deze studie dat er minder sekse-bias optreedt wanneer cognitieve vaardigheidsproeven en videosimulaties samen deel uitmaken van een testbatterij (zie ook Schmitt, Rogers, Chan, Sheppard & Jennings, 1997). Ten vierde willen wij ervoor pleiten om in het licht van het 'strengere' selectiebeleid van de Vlaamse Gemeenschap (Ministerie van de Vlaamse Gemeenschap, 1996) aandacht te besteden aan het 'verzwaren' van het casusgedeelte. Concreet kan men bijvoorbeeld overwegen om de kandidaten bij de 'Les' geen kopieën van de transparanten meer te geven. Verder kunnen vragen met slechts twee antwoordopties beter vermeden worden om onder meer het gokken te ontmoedigen. Ook de 'Stilleestekst' kan moeilijker gemaakt worden. Zo zou men kunnen beslissen dat de kandidaten bij het beantwoorden van de vragen hun tekst niet meer mogen raadplegen.

Aangezien deze studie slechts een eerste peiling is, moeten bovenstaande beleidsimplicaties met de nodige voorzichtigheid geïnterpreteerd worden. Het is dus belangrijk om het evaluatieonderzoek over het toelatingsexamen verder door te zetten, zowel longitudinaal als bij nieuwe groepen studenten.

Literatuurlijst

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (AERA, APA &

- NCME). (1998). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Baneke, J.J. (1989). Selectie van artsen. In het bijzonder vóór of tijdens de studie geneeskunde. *Bulletin Medisch Onderwijs*, 2, 42-49.
- Bartlett, C.J., Bobko, P., Mosier, S.B. & Hannon, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative for differential analysis. *Personnel Psychology*, 31, 233-241.
- Bobko, P. (1995). *Correlation and regression: Principles and applications for industrial/organizational psychology*. New York: McGraw-Hill.
- Cate, T.J. ten, Rossum, H.J.M. van & Vleuten, C.P.M. van der (1997). Belangen bij de selectie van de artsopleiding: Commentaar bij het rapport van de commissie Drenth. *Bulletin Medisch Onderwijs*, 16, 81-84.
- Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Coetsier, P. & Lievens, F. (1997). *Een andere kijk op evaluatie en selectie van kandidaat medische studenten: Ontwikkeling van op video opgenomen simulaties*. Congresbundel Onderwijsresearchdagen 1997 (pp. 191-194). Leuven: Katholieke Universiteit Leuven, Afdeling Didactiek.
- Glaser, K., Hojat, M., Veloski, J.J., Blacklow, R.S. & Goepp, E.C. (1992). Science, verbal or quantitative skills: Which is the most important predictor of physician competence? *Educational and Psychological Measurement*, 52, 395-406.
- Green, A., Peters, T.J. & Webster, D.J.T. (1993). Pre-clinical progress in relation to personality and academic profiles. *Medical Education*, 27, 137-142.
- Green, A., Peters, T.J. & Webster, J.T. (1991). An assessment of academic performance and personality. *Medical Education*, 25, 343-348.
- Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport (Conn.): Praeger.
- Lievens, F. & Coetsier, P. (1998). Een andere kijk op selectie van kandidaat medische studenten: Ontwikkeling van op video opgenomen simulaties. *Tijdschrift voor Hoger Onderwijs*, 16, 117-131.
- Lipton, A., Huxham, G.J. & Hamilton, D. (1984). Predictors of success in a cohort of medical students. *Medical Education*, 18, 203.
- McManus, I.C. & Richards, P. (1986). Prospective survey of performance of medical students du-

ring preclinical years. *British Medical Journal*, 293, 124-12.

Ministerie van de Vlaamse Gemeenschap (1996). Decreet houdende wijziging van het decreet van 12 juni 1991 betreffende de universiteiten in de Vlaamse Gemeenschap. *Belgisch Staatsblad* 19-09-96.

Minnaert, A. (1996). *Academic performance, cognition, metacognition and motivation. Assessing freshmen characteristics on task: A validation and replication study in higher education*. Niet gepubliceerd doctoraatsproefschrift, Katholieke Universiteit Leuven, België.

Mitchell, K., Haynes, R. & Koenig, J. (1994). Assessing the validity of the updated medical college admission test. *Academic Medicine*, 69 (5), 394-401.

Montague, W. & Odds, F.C (1990). Academic selection criteria and subsequent performance. *Medical Education*, 24, 44-47.

Motowidlo, S.J. & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, 66, 337-344.

Patel, V.L., Groen, G.J. & Frederiksen, C.H. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education*, 20 (1), 3-9.

Ree, M.J., Carretta, T.R., Earles, J.A. & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology*, 79, 298-301.

Roessler, R., Lester, J.W., Butler, W.T., Rankin, B. & Collins, F. (1978). Cognitive and non-cognitive variables in the prediction of preclinical performance. *Journal of Medical Education*, 53, 678-681.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L. & Jennings, D. (1997). Adverse impact and predictive efficiency using various predictor combinations. *Journal of Applied Psychology*, 82, 719-730.

Smal, J.A. (1995). Selectie met gewogen loting. *Bulletin Medisch Onderwijs*, 14, 97-101.

Society for Industrial and Organizational Psychology. (1987). *Principles for the validation and use of personnel selection procedures*. College Park, MD: Author.

Stinissen, J. & Steene, G. vander(1980). Oriëntering en selectie voor de universiteit. In *Gedrag, dynamische relatie en betekeniswereld. Liber Amicorum Prof. J.R. Nuttin*. Leuven: Universitaire Pers Leuven.

Thorndike, R.L. (1949). *Personnel selection: Test and measurement techniques*. New York: Wiley.

Tomlinson, R.W.S., Clack G.B., Pettingale, K.W., Anderson, J. & Ryan, K.C. (1977). The relative role of 'A' level chemistry, physics and biology in the medical course. *Medical Education*, 11, 103-108.

Vu, N.V., Dawson-Saunders, B. & Barrows, H.S. (1987). Use of Medical Reasoning Aptitude Test to help predict performance in medical school. *Journal of Medical Education*, 62, 325-335.

Walton, H.J. (1987). Personality assessment of future doctors. *Journal of the Royal Society of Medicine*, 80, 27-30.

Wilson, S., Suddick, R.P., Shay, J.S. & Hustmyer, F.E. (1981). Correlation of scores on embedded figures and mirror tracing with preclinical technique grades and PMAT scores of dental students. *Perceptual and Motor Skills*, 53 (1), 31-35.

Manuscript aanvaard: 29 oktober 2000

Auteurs

Filip Lievens is als docent aan de Universiteit Gent verbonden aan de Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie, Universiteit Gent. Hij is ook postdoctoraal onderzoeker bij het Fonds voor Wetenschappelijk Onderzoek- Vlaanderen (F.W.O.). Correspondentie-adres: Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie, Faculteit PPW, Universiteit Gent, Henri Dunantlaan 2, Gent B-9000, België. E-mail: [filip.lievens@rug.ac.be]

Pol Coetsier is als gewoon hoogleraar aan de Universiteit Gent verbonden aan de Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie, Universiteit Gent. Correspondentie-adres: Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie, Faculteit PPW, Universiteit Gent, Henri Dunantlaan 2, Gent B-9000, België. E-mail: [pol.coetsier@rug.ac.be]

Piet J. Janssen is als emeritus gewoon hoogleraar aan de Katholieke Universiteit Leuven verbonden aan de Afdeling Psychodiagnostiek en Psychologische begeleiding, Katholieke Universiteit Leuven. Correspondentie-adres: Afdeling Psychodiagnostiek en Psychologische begeleiding, Centrum voor Schoolpsychologie Faculteit PPW, Universiteit Leuven, Tiensestraat 102, Leuven B-3000, België. E-mail: [Piet.Janssen@psy.kuleuven.ac.be]

Christoph Decaestecker was ten tijde van het onderzoek als assistent verbonden aan de Vakgroep Personeelsbeleid en Arbeids- en Organisationspsychologie, Universiteit Gent.

Correspondentie-adres: Vakgroep Personeelsbeleid en Arbeids- en Organisationspsychologie, Faculteit PPW, Universiteit Gent, Henri Dunantlaan 2, Gent B-9000, België.

E-mail: [Christoph.Decaestecker@rug.ac.be]

Abstract

Predictive validity and gender bias of the Admission Exam ‘Medical and Dental Studies’ in Flanders: A first evaluation”

Since 1997 the system of open admission to medical and dental studies in Flanders has been given up. Yet, no evaluation of the Flemish Admission Exam has occurred. Therefore, this study evaluates the predictive validity and the gender bias of this exam. As part of the exam 941 candidates (359 men, 582 women; mean age = 18 years and 3 months) participated in four cognitive ability tests and four situational tests. Students' results at the end of the first year of medical and dental studies served as criterion measures. The correlation between the aggregate Admission Exam score and the first-year score was .35. The 'Reasoning Test' emerged as best predictor. Although differences were found between boys and girls in terms of their mean scores on some of the tests, these differences balanced out in the aggregate Admission Exam score. The Admission Exam did also not suffer from differential prediction.