

De predictieve validiteit van cognitieve en niet-cognitieve voorspellers van het Toelatings-examen "Arts en Tandarts" in Vlaanderen

F. Lievens en T. Buyse

Samenvatting

Deze studie is de eerste grootschalige studie naar de validiteit van het Vlaams toelatingsexamen "Arts en Tandarts". De steekproef bestaat uit alle studenten die hebben deelgenomen aan het toelatingsexamen tussen 1999 en 2003 en die vervolgens de opleiding geneeskunde zijn begonnen aan één van de Vlaamse universiteiten. De gevonden resultaten bevestigen dat cognitieve voorspellers valide instrumenten zijn bij selectie van studenten voor het medisch onderwijs. Een videoproef over interpersoonlijke vaardigheden is valide in latere jaren van de opleiding. Bovendien is deze videoproef een valide voorspeller in curricula waarbij vakken over interpersoonlijke vaardigheden ook meespelen om de studieresultaten te bepalen. Dit toont dat videoproeven een interessante en nuttige aanvulling kunnen vormen bij traditionele cognitieve voorspellers in selectie voor het hoger onderwijs.

1 Inleiding

In vele landen fungeren onderwijsprestaties en cognitief georiënteerde tests als de belangrijkste instrumenten om studenten al dan niet toe te laten tot het hoger onderwijs. Voorbeelden van cognitieve toetsen zijn intelligentietests, vorderingstoetsen en examens. Het is echter duidelijk dat deze cognitief georiënteerde voorspellers een belangrijk maar onvolledig beeld schetsen van de capaciteiten van studenten. Daarom bestaat reeds decennialang een belangrijke doelstelling van selectie- en onderwijsdeskundigen erin om deze cognitief georiënteerde voorspellers uit te breiden met andere voorspellers (Sackett, Schmitt, Ellingson, & Kabin, 2001; Ten Cate, Van Rossum, & Van der Vleuten, 1997). Hierin kaderen bijvoorbeeld het gebruik van interviews, persoonlijkheidsvragenlijsten en

videoproeven die peilen naar interpersoonlijke vaardigheden. De zoektocht naar "andere" voorspellers vloeit ook voort uit het feit dat universiteiten succesvolle prestaties van studenten steeds vaker breed definiëren. Dit betekent dat niet enkel de behaalde studieresultaten, maar ook bijvoorbeeld sociale en interpersoonlijke vaardigheden meespelen in de omschrijving van curricula.

Dezelfde aandachtspunten speelden een rol bij de ontwikkeling van het Vlaams toelatingsexamen "Arts en Tandarts". In Vlaanderen moeten studenten die de opleidingen geneeskunde of tandheelkunde willen aanvatten, slagen voor een toelatingsexamen. Net zoals vele examens in andere landen, peilt dit examen naar cognitief georiënteerde persoonsaspecten via een redeneertest en toetsen over vier wetenschapsvakken (biologie, fysica, scheikunde en wiskunde). Daarnaast omvat dit toelatingsexamen twee simulaties. Zo dienen kandidaten vragen te beantwoorden over een medische tekst en over een op video opgenomen gesprek tussen een arts en een patiënt. Deze videoproef peilt naar sociale en interpersoonlijke vaardigheden.

Tot op heden werd de validiteit van dit Vlaams toelatingsexamen "Arts en Tandarts" enkel onderzocht voor het eerste academiejaar voor een algemeen criterium, namelijk de studieresultaten in het eerste academiejaar (Lievens & Coetsier, 2002; Lievens, Coetsier, & Buyse, 2001; Lievens, Coetsier, Janssen, & Decaesteker, 2001). Daarom is de doelstelling van deze studie tweevoudig. Allereerst beogen wij de voorspellende kracht van dit Vlaams toelatingsexamen na te gaan voor de eerste vier studie jaren van de opleiding geneeskunde en tandheelkunde. Ten tweede willen wij de validiteit nagaan per curriculum. De universiteiten in Vlaanderen verschillen in deze nogal van elkaar. In sommige universiteiten hebben wetenschaps- en medische vakken een doorslaggevend gewicht bij het bepalen van de studieresultaten. Presta-

ties op interpersoonlijk georiënteerde vakken spelen formeel geen rol van betekenis bij het berekenen van de studieresultaten. In andere universiteiten worden zowel uitslagen op wetenschaps- en medische vakken als op vakken over interpersoonlijke vaardigheden gebruikt bij het berekenen van de studieresultaten.

2 Achtergrond

Zoals reeds gesteld, staan in vele toelatingsexamens cognitief georiënteerde voorspellers centraal. In de Verenigde Staten worden bijvoorbeeld de Scholastic Assessment Test (SAT), de American College Test (ACT) en de Graduate Record Examinations (GRE) veelvuldig gebruikt bij selectie voor het hoger onderwijs. Een recente meta-analyse van Kuncel, Hezlett en Ones (2001) bevestigde het belang van deze cognitieve voorspellers. Deze meta-analyse toonde aan dat een combinatie van algemene cognitieve vaardigheid en vakspecifieke toetsen (bijvoorbeeld wetenschapstoetsen) de hoogste validiteit had bij het voorspellen van academisch succes. Gelijkaardige resultaten werden gevonden bij de selectie van medische studenten (Green, Peters, & Webster, 1991; Minnaert, 1996; Mitchell, Haynes, & Koenig, 1994; Montague & Odds, 1990; Powis, 1994; Roessler, Lester, Butler, Rankin, & Collins, 1978; Tomlinson, Clack, Pettingale, Anderson, & Ryan, 1977; Vu, Dawson-Saunders, & Barrows, 1987). Daarom formuleren we de volgende hypothese:

Hypothese 1: De cognitief georiënteerde toetsen van het toelatingsexamen (wetenschapstoetsen en redeneertest) zullen significante voorspellers zijn voor het studieresultaat in de verschillende jaren van de medische opleiding.

Naast cognitief georiënteerde toetsen fungeren ook vaak "alternatieve" proeven als basis voor selectiebeslissingen in het hoger onderwijs. Het woord *alternatief* geeft aan dat deze proeven eerder niet-cognitieve factoren meten. De onderliggende idee is dat deze proeven een breder scala aan vaardigheden meten. Bij het Vlaams toelatingsexamen werd in dit kader gekozen voor een video-proef die interpersoonlijke vaardigheden

meet. Dergelijke video-proeven, ook wel 'video-based situational judgment tests' genoemd, worden de laatste tijd steeds populairder bij de selectie van personeel in een organisatiecontext. Bij zulke proeven worden kandidaten geconfronteerd met hypothetische, functie-relevante situaties (ofwel op papier geschreven ofwel op video), waarna zij de passende reactie dienen te kiezen uit een reeks antwoordalternatieven (Motowidlo, Dunnette, & Carter, 1990; Weekley & Jones, 1999). Wetenschappelijk onderzoek laat ook positieve resultaten zien bij het gebruik van video-proeven. Zij hebben een goede voorspellende kracht (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001), verklaren extra variantie bovenop de validiteit van cognitieve- en persoonlijkheidstests (Chan & Schmitt, 2002; Clewenger, Pereira, Weichmann, Schmitt, & Schmidt-Harvey, 2001) en lokken positieve reacties uit bij kandidaten ('face validity'). Nadeel is dat de ontwikkeling van video-proeven een dure en tijdsintensieve aangelegenheid is.

Hoewel video-proeven over interpersoonlijke vaardigheden reeds hun waarde hebben bewezen in een medische trainingscontext en onderwijscontext (Bögels & Kreutzkamp, 1990; Van der Molen & Vermeulen, 1995; Smit & Van der Molen, 1996), zijn wij niet op de hoogte van studies die hun effectiviteit aantonen bij de selectie van studenten voor het hoger onderwijs. Dit is verrassend, omdat bovengenoemd onderzoek bij de selectie van personeel laat zien dat video-proeven een goede voorspellende kracht hebben en toelaten om andere constructen te meten dan intelligentie. Een belangrijke kanttekening is hier op zijn plaats. Van een video-proef die interpersoonlijke vaardigheden meet, wordt theoretisch niet verwacht dat deze een cognitief geladen criterium voorspelt. Daarentegen zal een interpersoonlijk georiënteerde video-proef een goede voorspeller zijn voor academisch succes in een curriculum met interpersoonlijke vakken. Ook zal een dergelijke video-proef de resultaten die studenten behalen op interpersoonlijke vakken voorspellen. Verwachtingen over de validiteit van een video-proef vereisen dus de nodige aandacht voor het criterium. In Vlaanderen is het mogelijk dit te toetsen, omdat de universiteiten

nogal verschillen wat betreft het aangeboden curriculum. In universiteiten waar studieresultaten formeel enkel gebaseerd zijn op prestaties op wetenschaps- en medische vakken zullen de cognitief georiënteerde tests van het toelatingsexamen een voorspellende kracht hebben voor de studieresultaten. In universiteiten waar prestaties van studenten op vakken over interpersoonlijke vaardigheden ook de studieresultaten meebepalen, kunnen we veronderstellen dat ook de interpersoonlijke videoproef een goede voorspellende kracht zal hebben. Dit geeft aanleiding tot de volgende hypothesen:

Hypothese 2: Als de algemene studieresultaten in de verschillende jaren van de medische opleiding niet enkel gebaseerd zijn op wetenschaps- en medische vakken maar ook op interpersoonlijke vakken, zal een videoproef die interpersoonlijke vaardigheden meet, een significante voorspeller zijn voor deze studieresultaten.

Hypothese 3: Als de algemene studieresultaten in de verschillende jaren van de medische opleiding niet enkel gebaseerd zijn op wetenschaps- en medische vakken, maar ook op interpersoonlijke vakken, zal een videoproef die interpersoonlijke vaardigheden meet, een bijkomend stuk van de variantie verklaren (bovenop de variantie verklaard door de cognitief georiënteerde voorspellers).

Hypothese 4: Een videoproef die interpersoonlijke vaardigheden meet, zal een significante voorspeller zijn voor de studieresultaten die studenten behalen op vakken over interpersoonlijke vaardigheden.

3 Methode

3.1 Steekproef

De totale steekproef bestond uit 7197 kandidaten (2606 mannen en 4591 vrouwen) die deelnamen aan het toelatingsexamen “Arts en Tandarts” in Vlaanderen tussen 1999 en 2003. De gemiddelde leeftijd van de deelnemers was 18 jaar en 11 maanden. Over het algemeen was het slagingspercentage voor het toelatingsexamen ongeveer 30%.

Enkel deelnemers die slaagden voor het examen, daaropvolgend met de opleiding begonnen en deze verder doorliepen, werden in

het onderzoek opgenomen. In totaal beschikten we over de studieresultaten van 1768 studenten in het eerste jaar, 1087 studenten in het tweede jaar, 676 studenten in het derde jaar en 305 studenten in het vierde jaar. Dit geringe aantal studenten in de latere jaren van het curriculum is slechts deels te wijten aan studenten die afhaken. Hoofdzakelijk is het geringe aantal studenten in latere jaren een gevolg van het nog niet voorhanden zijn van betreffende criteriumdata op het moment van deze studie. Zo hadden wij studieresultaten voor de eerste vier academiejaren van studenten die in 1999 slaagden voor het toelatingsexamen, studieresultaten voor de eerste drie academiejaren van studenten die in 2000 slaagden, etc. Voor de cohort van 2003 waren er nog geen studieresultaten beschikbaar.

3.2 Voorspellers

De voorspellers waren de verschillende onderdelen van het toelatingsexamen. Ieder jaar werd het examen op één dag georganiseerd in een grote zaal te Brussel. In de voormiddag legden de kandidaten de vier wetenschapstoetsen af. In de namiddag volgden een redeneertest, een stillestekst en een videoproef (op video opgenomen gespreksfragmenten tussen een arts en een patiënt).

Wetenschapstoetsen. Ieder jaar werden vier wetenschapstoetsen (biologie, fysica, scheikunde en wiskunde) ontwikkeld voor het toelatingsexamen. Iedere toets bestond uit 10 vragen met telkens 4 antwoordalternatieven. De kandidaten hadden 180 minuten om deze 40 vragen op te lossen. Iedere toets werd onder leiding van een hoogleraar (specialist in het betreffend vak) opgesteld.

Redeneertest. Deze test bestond uit 50 items met 5 antwoordalternatieven. De items werden ieder jaar willekeurig gekozen uit een grote itemverzameling. De test bestond uit items met een in moeilijkheidsgraad opklimmende serie van verbale, numerieke en diagrammatische opdrachten. De tijdslimiet was 50 minuten. Eerder onderzoek bevestigde de goede betrouwbaarheid en voorspellende kracht van deze test bij medische studenten. Minnaert (1996) rapporteerde een interne consistentie van 0.84 en een validiteitscoëfficiënt van 0.36 bij het voorspellen van studie-

resultaten in het eerste jaar geneeskunde of tandheelkunde.

Stillestekst. Deze proef werd ontwikkeld voor het toelatingsexamen. Het achterliggende idee was dat kandidaten in staat moeten zijn om een eenvoudige tekst over een medisch onderwerp te lezen en te begrijpen. Voorbeelden van onderwerpen waren diabetes, rugpijn, etc. De tekst was ongeveer 10 pagina's lang en werd opgebouwd als een wetenschappelijk artikel inclusief tabellen en figuren. Statistische analyses werden vermeden en alle technische termen werden verklaard in een bijgevoegde woordenlijst. Studenten hadden 50 minuten om de tekst te lezen en de 30 vragen op te lossen. Alle vragen waren 'multiple choice'-vragen met vier antwoordalternatieven. Elk jaar werd dezelfde procedure gevolgd om de tekst en de bijbehorende vragen op te stellen. Een bestaand artikel in een medisch handboek of tijdschrift fungeerde meestal als uitgangspunt. Een hoogleraar in de geneeskunde of tandheelkunde breidde vervolgens de tekst uit. De vragen en antwoordalternatieven werden opgesteld in overleg met twee hoogleraren geneeskunde.

Videotest. Ook deze proef werd ontwikkeld voor het toelatingsexamen. Via de videoproef beoogden wij te peilen naar interpersoonlijke en communicatieve vaardigheden. De videoproef bestond uit een aantal korte, op video opgenomen fragmenten van interpersoonlijke voorvallen bij een interactie tussen een arts en een patiënt. Elk jaar handelde de video over één specifiek thema (bijvoorbeeld een patiënt met misselijkheid, borstpijn, etc.). Hoewel het thema elk jaar veranderde, keerden de gebruikte kritische incidenten terug (meedelen van slecht nieuws, reageren op klachten van de patiënt over de behandeling, etc.). De kritische incidenten waren gebaseerd op gesprekken met hoogleraren in de geneeskunde en ervaren huisartsen. Ieder jaar volgden wij ook hetzelfde stramien bij de ontwikkeling van deze proef (Motowidlo et al., 1990; Weekley & Jones, 1997). Er werd een script geschreven waarin de kritische incidenten werden verwerkt. Twee hoogleraren huisartsgeneeskunde lazen de scripts na op hun realiteitsgehalte. Elk jaar werden acteurs ingehuurd die de

scripts naspeelden in een opnamestudio. Een ervaren huisarts was hierbij aanwezig. In totaal werden 30 multiple-choice-vragen opgesteld met telkens vier antwoordalternatieven. Een groep experts (ervaren huisartsen en hoogleraren in de geneeskunde) bepaalde bij iedere vraag het juiste antwoord. Voor een uitvoerige beschrijving van de ontwikkeling van de videoproef verwijzen we naar Lievens en Coetsier (1998). Omwille van de testveiligheid van deze proeven kunnen hier geen voorbeelditems gepresenteerd worden.

Totaalscore toelatingsexamen. Deze totaalscore werd gebruikt om te beslissen wie al dan niet slaagde voor het examen, en was een gewogen som van elk van voorgaande predictoren. Daarna werd een minimale 'cut-off'-score bepaald voor deze totaalscore. Het gewicht van elke predictor en de cut-off-score werden bij wet bepaald.

3.3 Criterium

De studieresultaten van studenten in de eerste vier studiejaar van de medische opleiding werden opgevraagd bij alle Vlaamse universiteiten. In Vlaanderen worden de studieresultaten uitgedrukt op een schaal van 0 tot 20, waarbij een hogere score betere resultaten aangeeft. Studieresultaten in de verschillende jaren correleerden hoog met elkaar (van .73 tot .78). Deze correlaties zijn gelijklopend met de resultaten van een recente meta-analyse over de betrouwbaarheid (in de vorm van temporele stabiliteit) van studieresultaten (bijv. 'grade point average') (Vey et al., 2003). Wij beschikten ook over de studieresultaten van de studenten op de verschillende vakken. De prestaties van studenten op de verschillende vakken worden eveneens gemeten op een schaal van 0 tot 20, waarbij hogere scores betere resultaten aangeven.

4 Resultaten

4.1 Voorbereidende analyses

Aangezien we onze hypothesen toetsten op gegevens die verzameld werden over verschillende jaren (van 1999 tot 2003), gingen we eerst na of de onderliggende structuur van het toelatingsexamen gelijk was over deze

Tabel 1

Tests van invariantie voor 'multiple'-groep-driefactormodel van de scores op het toelatingsexamen over examenjaren heen

	χ^2	$\Delta \chi^2$	RNI	CFI	ΔCFI	AGFI	RMSEA	90% Betrouw.- interv. RMSEA
Gelijk aantal factoren	467.81**	--	.918	.949	--	.953	.032	[.029 - .035]
Gelijke factorladingen	469.40**	1.59	.923	.950	-.001	.955	.031	[.028 - .034]
Gelijke foutvarianties	471.18**	1.78	.929	.950	.000	.958	.030	[.027 - .032]
Gelijke factor varianties/covarianties	471.90**	.72	.935	.951	-.00	.961	.029	[.026 - .031]

Noot. $N = 6005$. Deze N verschilt van de totale groep deelnemers ($N = 7185$, zie Tabel 2), omdat deze analyses dienden uitgevoerd te worden met zogenaamde 'listwise deletion of cases'.

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

jaren heen. Meer specifiek toetsten wij of aan de vereiste criteria in verband met meetequivalentie voldaan was (Hoyle & Smith, 1994). Zoals Tabel 1 aangeeft, toonde een theoretisch model met drie factoren, namelijk een cognitieve factor (bestaande uit de redeneertest en de vier wetenschapstoetsen, zie Kuncel et al., 2001), een factor waarop de stilleestekst laadde en een factor gerelateerd aan de score op de videoproef, een goede passing met de empirische gegevens ($RNI = .935$, $CFI = .951$ en $RMSEA = .029$). Dit model bleek ook invariant over de jaren, aangezien factorvorm, factorladingen, foutvarianties en factorvarianties/covarianties gelijklopend waren. Op basis hiervan rapporteren we bij de verdere analyses de resultaten voor deze drie factoren: het Cognitief totaal, de Stilleestekst en de Videoproef.

Hoewel het meetmodel invariant was over de jaren heen, kunnen de gemiddelde scores op de tests en proeven toch verschillen van jaar tot jaar door verschillen in moeilijkheidsgraad. Daarom standaardiseerden we de scores van de studenten per examenjaar en per test/proef. Hetzelfde gold voor de criteriumgegevens. We standaardiseerden de algemene studieresultaten en de studieresultaten op de vakken van de studenten per academiejaar en per universiteit.

4.2 Analyses op niveau van algemene studieresultaten

Tabel 2 toont de gemiddelden, standaarddeviaties en onderlinge correlaties van de voorspellers. Deze tabel is gebaseerd op alle kandidaten die het toelatingsexamen aflegden tussen 1999 en 2003.

In Tabel 2 is te zien dat de correlaties tussen de scores op de drie testtypes klein tot middelmatig waren. De correlatie tussen de cognitieve factor en de videoproef was 0.19, wat aantoont dat de videoproef weinig cognitief geladen is. De meta-analyse van McDaniel e.a. (2001) vond een gemiddelde correlatie van 0.36 tussen cognitieve vaardigheid en dergelijke proeven.

Tabel 2

Gemiddelden (M), standaarddeviaties (SD) en onderlinge correlaties (r) van voorspellers in de totale groep ($N = 7185$)

	M	SD	r		
			1	2	3
1 Cognitief totaal	11.35	2.69	--		
2 Stilleestekst	15.81	4.87	.38**	--	
3 Videoproef	18.73	3.11	.19**	.24**	--
4 Totaal	20.58	5.17	.92**	.49**	.29**

Noot. Alhoewel alle analyses uitgevoerd werden op gestandaardiseerde scores, toont deze tabel de ruwe scores over examens heen. De maximum score is telkens 30, voor het totaal is dit 40.

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

Tabel 3 toont de gemiddelden, standaarddeviaties en onderlinge correlaties van voorspellers en criteria. Deze tabel is dus enkel gebaseerd op kandidaten die slaagden voor het toelatingsexamen en die daarna aan hun opleiding geneeskunde of tandheelkunde begonnen. Omdat er zowel sprake was van indirecte spreidingsbreedte (studenten werden geselecteerd op grond van hun totaalscore op het toelatingsexamen) als directe spreidingsbreedte (criteriumgegevens waren enkel beschikbaar voor studenten die slaagden op het

examen), pasten we de formules voor de correctie van multivariate spreidingsbreedte (zie Ree, Carretta, Earles, & Albert, 1994) toe op de ongecorrigeerde correlatiematrix. Pas nadat deze correcties werden uitgevoerd, corrigeerden we de correlaties ook voor onbetrouwbaarheid van het criterium (studieresultaten). Om een schatting te bekomen voor de (on)betrouwbaarheid van het criterium, gebruikten we de gemiddelde correlatie (.75) tussen de studieresultaten in opeenvolgende jaren (zie Tabel 3). Zoals reeds gesteld, komt deze waarde overeen met de waarde voor de

betrouwbaarheid van studieresultaten zoals bekomen in de meta-analyse door Vey e.a. (2003). Statistische significantie van de correlaties werd bepaald vóór de correctie werd toegepast (Sackett & Yang, 2000). De ongecorrigeerde correlaties staan onder de diagonaal, de gecorrigeerde correlaties staan boven de diagonaal.

Tabel 3 vormt de basis voor het toetsen van Hypothese 1, namelijk dat cognitief georiënteerde predictoren goede voorspellers zouden zijn voor de studieresultaten behaald aan de universiteit. De resultaten toonden aan

Tabel 3

Gemiddelden (*M*), standaarddeviaties (*SD*) en onderlinge correlaties (*r*) tussen voorspellers en criteria

	<i>M</i>	<i>SD</i>	<i>r</i>				
			1	2	3	4	5
<i>Jaar 1 (N = 1768)</i>							
1 Cognitief totaal	14.13	1.66	--	.17	.04	.88	.52
2 Stilleestekst	16.94	4.50	.01	--	.14	.31	.12
3 Videoproef	19.23	2.87	-.05*	.12**	--	.16	.08
4 Totaalscore toelatingsexamen	25.09	3.78	.77**	.21**	.10**	--	.51
5 Studieresultaat jaar 1	13.27	2.53	.33**	.03	.03	.31**	--
<i>Jaar 2 (N = 1087)</i>							
1 Cognitief totaal	13.88	1.64	--	.14	.01	.87	.46
2 Stilleestekst	15.62	4.23	.03	--	.17	.26	.09
3 Videoproef	19.60	2.87	-.06*	.15**	--	.14	.09
4 Totaalscore toelatingsexamen	24.84	4.03	.77**	.18**	.09**	--	.45
5 Studieresultaat jaar 2	14.07	1.98	.30**	.03	.05	.28**	--
<i>Jaar 3 (N = 676)</i>							
1 Cognitief totaal	13.56	1.63	--	.37	.14	.97	.55
2 Stilleestekst	16.93	4.01	.10*	--	.26	.48	.29
3 Videoproef	19.98	2.98	-.03	.20**	--	.24	.20
4 Totaalscore toelatingsexamen	25.99	2.29	.91**	.30**	.14**	--	.55
5 Studieresultaat jaar 3	14.21	1.60	.29**	.11**	.11**	.30**	--
<i>Jaar 4 (N = 305)</i>							
1 Cognitief totaal	13.51	1.49	--	.52	.45	.96	.37
2 Stilleestekst	18.56	4.25	.31**	--	.53	.56	.27
3 Videoproef	21.35	3.09	.26**	.44**	--	.48	.35
4 Totaalscore toelatingsexamen	26.21	2.27	.91**	.38**	.32**	--	.36
5 Studieresultaat jaar 4	14.45	1.80	.21**	.14*	.24**	.20**	--

Noot. Alhoewel alle analyses uitgevoerd werden op gestandaardiseerde scores, toont deze tabel de ruwe scores. De maximum score was telkens 30, met uitzondering van de totaalscore op het toelatingsexamen (maximum score = 40) en het studieresultaat tijdens de opleiding (maximum score = 20). De ongecorrigeerde correlaties staan onder de diagonaal, de gecorrigeerde correlaties staan boven de diagonaal. Correlaties werden gecorrigeerd voor multivariate spreidingsbreedte en onbetrouwbaarheid in het criterium. Statistische significantie werd bepaald vóór de correcties werden toegepast.

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

dat een combinatie van alle cognitieve toetsen van het examen significant en consistent correleerde met de studieresultaten van de studenten in de eerste vier studie jaren van de opleiding. Dit is het meest duidelijk voor het eerste studiejaar. In dit eerste jaar bedroeg de gecorrigeerde correlatie tussen de cognitieve voorspellers en de studieresultaten 0.52. Ook in de andere jaren is er een hoge correlatie tussen de cognitieve voorspeller en de studieresultaten, alhoewel deze wel ietwat afneemt in het laatste jaar (.37). Deze resultaten bevestigen Hypothese 1. Aangezien de andere hypothesen betrekking hebben op de verschillende curricula, gaan we nu over op afzonderlijke analyses per type curriculum.

4.3 Analyses op curriculumniveau

Onze andere hypothesen stellen dat een goede voorspellende kracht van een proef die interpersoonlijke vaardigheden meet (zoals de videoproef van het toelatingsexamen) enkel verwacht kan worden indien het criterium ook een niet-cognitieve dimensie omvat. Meer bepaald, stelt Hypothese 2 dat een videoproef (die interpersoonlijke vaardigheden meet) de studieresultaten significant kan voorspellen indien deze laatste niet enkel gebaseerd zijn op wetenschaps- en medische vakken, maar eveneens op interpersoonlijke vakken.

Om deze hypothese te toetsen, bekeken we het curriculum van de vier universiteiten in onze steekproef. In totaal werden in deze vier universiteiten en in de eerste vier jaren van de opleiding 105 vakken gedoceerd. De twee auteurs gingen de inhoud van de vakken na, en gaven ieder vak een beoordeling, variërend van 1 = *bij dit vak ligt de nadruk helemaal niet op interpersoonlijke/communicatieve vaardigheden bij arts-patiëntinteracties* tot 5 = *bij dit vak ligt de nadruk heel sterk op interpersoonlijke/communicatieve vaardigheden bij arts-patiëntinteracties*. Overeenstemming tussen de beoordelingen bedroeg 0.92 ('intra-class correlation' 2.1, Shrout & Fleiss, 1979). Discrepanties werden nadien gemakkelijk opgelost door overleg tussen de twee auteurs.

Vakken (zoals klinische en communicatieve vaardigheden, communicatie, etc.) die een beoordeling kregen van 3 of hoger werden

beschouwd als vakken met een nadruk op interpersoonlijke en communicatieve vaardigheden. Opvallend was dat de curricula van de betrokken universiteiten duidelijk van elkaar onderscheiden konden worden in termen van de hoeveelheid aandacht die aan dit soort vakken werd besteed gedurende de eerste vier jaren van de opleiding. Twee types curricula kwamen naar voren. In het curriculum van twee universiteiten werden de studieresultaten deels bepaald door interpersoonlijk georiënteerde vakken. Het gewicht van de interpersoonlijke vakken in het berekenen van de studieresultaten in het eerste, tweede, derde en vierde jaar was hier respectievelijk 0.05, 0.11, 0.22 en 0.27. In het curriculum van de twee andere universiteiten hadden interpersoonlijke vakken veel minder belang in het bepalen van de studieresultaten. Het gewicht van de interpersoonlijke vakken in het berekenen van de studieresultaten in het eerste, tweede, derde en vierde jaar in deze universiteiten was respectievelijk 0.00, 0.00, 0.05 en 0.10. Merk op dat dit niet betekent dat er geen interpersoonlijke vakken werden gedoceerd aan deze universiteiten. Integendeel, de studenten dienden ook in deze universiteiten te slagen voor interpersoonlijke vakken, alleen was het zo dat de prestaties van de studenten op deze interpersoonlijke vakken formeel niet meetelden (i.e., studenten moesten deze vakken wel volgen, maar de resultaten hadden weinig tot geen gewicht bij het berekenen van de studieresultaten).

Tabel 4 en Tabel 5 tonen de correlaties tussen voorspellers en criteria voor beide onderscheiden curricula. Zoals te zien is, had de videoproef lage gecorrigeerde correlaties met de studieresultaten in universiteiten waar weinig aandacht besteed wordt aan interpersoonlijke vakken (.03, .07, .01 en .20 in respectievelijk het eerste, tweede, derde en vierde jaar). De validiteit van de cognitieve voorspeller vertoont een omgekeerde trend. Deze validiteit is hoog in de eerste jaren en ietwat lager in latere jaren.

In universiteiten waar wel aandacht besteed wordt aan interpersoonlijke vakken, bleek de videoproef echter een significante voorspeller voor de studieresultaten in de eerste jaren van de opleiding. De gecorrigeerde validiteiten bedroegen 0.12, 0.14,

0.40 en 0.55 in respectievelijk het eerste, tweede, derde en vierde jaar. Voor de laatste twee jaar was het verschil tussen de gecorrigeerde correlaties in beide curricula statistisch significant. We kunnen dus Hypothese 2 bevestigen. Merk verder op dat de cognitieve voorspeller het in dit curriculum ook goed doet (ook in latere jaren).

Hypothese 3 stelde dat, indien de studie-resultaten niet enkel gebaseerd waren op cognitieve vakken, maar ook op interpersoonlijk

georiënteerde vakken, een videoproef die interpersoonlijke capaciteiten meet, variantie zou verklaren bovenop de variantie die verklaard wordt door de cognitieve voorspellers. Om deze hypothese te toetsen, voerden we een hiërarchische regressieanalyse uit per curriculumtype. We gebruikten de gecorrigeerde matrices (zie boven) als input bij de hiërarchische regressieanalyses. Statistische significantie werd bepaald vóór de correcties werden doorgevoerd (door de hiërarchische

Tabel 4

Gemiddelden (*M*), standaarddeviaties (*SD*) en onderlinge correlaties (*r*) tussen variabelen en criteria voor curricula met minimale aandacht voor interpersoonlijke vakken

	<i>M</i>	<i>SD</i>	<i>r</i>				
			1	2	3	4	5
<i>Jaar 1 (N = 822)</i>							
1 Cognitief totaal	14.26	1.67	--	.19	.02	.89	.56
2 Stilleestekst	16.99	4.55	.04	--	.07	.31	.14
3 Videoproef	19.18	2.81	-.06	.05	--	.12	.03
4 Totaalscore toelatingsexamen	25.23	3.82	.79**	.21**	.08*	--	.57
5 Studieresultaat jaar 1	13.08	2.90	.35**	.04	-.01	.35**	--
<i>Jaar 2 (N = 484)</i>							
1 Cognitief totaal	14.01	1.63	--	.17	-.01	.89	.46
2 Stilleestekst	15.58	4.24	.06	--	.10	.27	.15
3 Videoproef	19.56	2.90	-.08	.09*	--	.11	.07
4 Totaalscore toelatingsexamen	25.08	3.99	.80**	.19**	.07	--	.47
5 Studieresultaat jaar 2	14.26	2.17	.28**	.07	.04	.29**	--
<i>Jaar 3 (N = 308)</i>							
1 Cognitief totaal	13.70	1.59	--	.35	-.05	.97	.46
2 Stilleestekst	16.95	3.98	.09	--	.16	.45	.23
3 Videoproef	20.02	2.98	-.14*	.16**	--	.05	.01
4 Totaalscore toelatingsexamen	26.18	2.25	.91**	.27**	.03	--	.47
5 Studieresultaat jaar 3	14.20	1.85	.23**	.08	.00	.24**	--
<i>Jaar 4 (N = 170)</i>							
1 Cognitief totaal	13.59	1.46	--	.46	.35	.96	.28
2 Stilleestekst	18.62	4.25	.27**	--	.45	.51	.16
3 Videoproef	21.59	2.89	.19*	.37**	--	.39	.20
4 Totaalscore toelatingsexamen	26.35	2.28	.91**	.34**	.26**	--	.28
5 Studieresultaat jaar 4	14.43	2.10	.15*	.07	.12	.15*	--

Noot. Alhoewel alle analyses uitgevoerd werden op gestandaardiseerde scores, toont deze tabel de ruwe scores. De maximum score was telkens 30, met uitzondering van de totaalscore op het toelatingsexamen (maximum score = 40) en het studieresultaat tijdens de opleiding (maximum score = 20). De ongecorrigeerde correlaties staan onder de diagonaal, de gecorrigeerde correlaties staan boven de diagonaal. Correlaties werden gecorrigeerd voor multivariate spreidingsbreedte en onbetrouwbaarheid in het criterium. Statistische significantie werd bepaald vóór de correcties werden toegepast.

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

Tabel 5

Gemiddelden (*M*), standaarddeviaties (*SD*) en onderlinge correlaties (*r*) tussen variabelen en criteria voor curricula met veel aandacht voor interpersoonlijke vakken

	<i>M</i>	<i>SD</i>	<i>r</i>				
			1	2	3	4	5
<i>Jaar 1 (N = 714)</i>							
1 Cognitief totaal	14.10	1.67	--	.15	.04	.86	.49
2 Stilleestekst	16.80	4.47	.00	--	.15	.30	.10
3 Videoproef	19.29	2.84	-.06	.12**	--	.17	.12
4 Totaalscore toelatingsexamen	24.94	3.86	.75**	.20**	.11**	--	.46
5 Studieresultaat jaar 1	13.36	2.10	.31**	.03	.07	.27**	--
<i>Jaar 2 (N = 433)</i>							
1 Cognitief totaal	13.88	1.68	--	.12	-.01	.83	.48
2 Stilleestekst	15.37	4.14	.02	--	.18	.24	.05
3 Videoproef	19.74	2.72	-.08	.16**	--	.14	.14
4 Totaalscore toelatingsexamen	24.61	4.23	.72**	.17**	.10*	--	.44
5 Studieresultaat jaar 2	13.86	1.78	.33**	.00	.10*	.28**	---
<i>Jaar 3 (N = 256)</i>							
1 Cognitief totaal	13.49	1.71	--	.40	.19	.97	.70
2 Stilleestekst	16.74	4.02	.12	--	.27	.52	.40
3 Videoproef	20.11	2.84	.00	.20**	--	.29	.40
4 Totaalscore toelatingsexamen	25.91	2.38	.91**	.33**	.18**	--	.72
5 Studieresultaat jaar 3	14.12	1.25	.40**	.17**	.27**	.42**	--
<i>Jaar 4 (N = 111)</i>							
1 Cognitief totaal	13.50	1.53	--	.57	.54	.97	.52
2 Stilleestekst	18.40	4.28	.35**	--	.59	.61	.44
3 Videoproef	21.08	3.11	.33**	.48**	--	.56	.55
4 Totaalscore toelatingsexamen	26.14	2.22	.91**	.42**	.38**	--	.52
5 Studieresultaat jaar 4	14.46	1.12	.30**	.25**	.38**	.29**	--

Noot. Alhoewel alle analyses uitgevoerd werden op gestandaardiseerde scores, toont deze tabel de ruwe scores. De maximum score was telkens 30, met uitzondering van de totaalscore op het toelatingsexamen (maximum score = 40) en het studieresultaat tijdens de opleiding (maximum score = 20). De ongecorrigeerde correlaties staan onder de diagonaal, de gecorrigeerde correlaties staan boven de diagonaal. Correlaties werden gecorrigeerd voor multivariate spreidingsbreedte en onbetrouwbaarheid in het criterium. Statistische significantie werd bepaald vóór de correcties werden toegepast.

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

regressie ook uit te voeren op de ongecorrigeerde correlatiematrix). Omdat cognitieve tests traditioneel worden gebruikt bij medische selectie, voerden wij deze groep tests in een eerste stap in. In een tweede stap werd de stilleestekst toegevoegd. Als laatste voegden we de videoproef toe aan het regressiemodel. De resultaten van de analyse bevinden zich in Tabel 6. De videoproef verklaarde geen bijkomende variantie in het curriculumtype

zonder aandacht voor interpersoonlijke vakken. Dit was wel het geval in het curriculumtype waar hieraan wel formeel aandacht besteed werd. De videoproef verklaarde in het tweede curriculumtype 1% bijkomende variantie in het eerste jaar, 2% in het tweede jaar, 6% in het derde jaar en 7% in het vierde jaar. Hiermee is Hypothese 3 bevestigd.

Zoals we reeds aangaven, waren op het moment van deze studie niet alle criterium-

Tabel 6

Samenvatting van hiërarchische regressieanalyses van voorspellers op totaalscore in de eerste vier jaren opgedeeld per type curriculum

	Curriculum met minimale aandacht voor interpersoonlijke vakken					Curriculum met veel aandacht voor interpersoonlijke vakken				
	β	t	p	R^2	ΔR^2	β	t	p	R^2	ΔR^2
Totaalscore jaar 1										
Cognitief totaal	.56	10.60	.00	.32	.32**	.48	8.98	.00	.24	.24**
Stilleestekst	.04	.80	.42	.32	.00	.02	.38	.70	.24	.00
Videoproef	.01	.25	.80	.32	.00	.10	2.47	.01	.25	.01
Totaalscore jaar 2										
Cognitief totaal	.45	6.42	.00	.21	.21**	.48	7.57	.00	.23	.23**
Stilleestekst	.06	1.19	.24	.22	.00	-.03	-.59	.56	.23	.00
Videoproef	.06	1.22	.22	.22	.00	.15	2.88	.00	.25	.02**
Totaalscore jaar 3										
Cognitief totaal	.44	4.02	.00	.21	.21**	.62	7.10	.00	.50	.50**
Stilleestekst	.07	1.03	.30	.22	.01	.08	1.31	.19	.51	.02*
Videoproef	.02	.37	.71	.22	.00	.26	4.53	.00	.58	.06**
Totaalscore jaar 4										
Cognitief totaal	.24	1.73	.08	.08	.08*	.30	1.91	.06	.27	.27**
Stilleestekst	-.01	-.10	.92	.08	.00	.06	.48	.64	.30	.03
Videoproef	.12	1.21	.23	.09	.01	.35	2.89	.00	.38	.07**

Noot. De gecorrigeerde matrices dienden als input voor de regressieanalyses. Statistische significantie werd bepaald vóór de correcties werden toegepast (door dezelfde regressieanalyses uit te voeren op de ongecorrigeerde matrices). Parameterschattingen zijn voor de laatste stap, niet voor de eerste. Door afronding verschilt de ΔR^2 0.01 van de cumulatieve R^2 .

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

gegevens beschikbaar. De gegevens in het derde en vierde jaar waren enkel beschikbaar voor studenten die in 1999 en 2000 voor het toelatingsexamen slaagden, terwijl de gegevens voor het eerste jaar beschikbaar waren voor de examens 1999, 2000, 2001 en 2002. Hierdoor is het mogelijk dat de toename van de voorspellende kracht van de videoproef over de jaren heen niet enkel een gevolg is van curriculumverschillen, maar ook van verschillen in de geselecteerde groep. Om deze reden voerden we alle analyses uit op de groep waarvan het grootste deel van de criteriumgegevens beschikbaar waren (de studenten die slaagden in 1999 en 2000). De resultaten toonden ook voor deze beperkte groep een stijgende trend aan in de voorspellende kracht van de videoproef.

4.4 Analyses op vakniveau

Hypothese 4 stelde dat een videoproef over interpersoonlijke vaardigheden een goede voorspeller zou zijn voor de studieresultaten op interpersoonlijke vakken. Om deze hypothese te toetsen, berekenden we validiteits-

coëfficiënten tussen de videoproef en de vakscores van studenten. Ter vergelijking berekenden wij ook de validiteitscoëfficiënten tussen het cognitief totaal en deze vakscores. Omdat er 105 vakken waren, gaf dit ons 105 coëfficiënten voor de videoproef en 105 voor het cognitief totaal. Daarna correleerden we deze coëfficiënten met de beoordeling die de vakken kregen op onze interpersoonlijke schaal (zie boven). De correlaties tussen de videoproefcoëfficiënten en de beoordelingen op de interpersoonlijke schaal waren positief ($r = .21$, $p < .05$). Dit duidt erop dat de videoproef globaal een betere voorspellende kracht heeft bij interpersoonlijk georiënteerde vakken. Hiermee is Hypothese 4 bevestigd.

5 Discussie

Deze studie is de eerste grootschalige studie naar de validiteit van het Vlaams toelatingsexamen “Arts en Tandarts”. Dit is ook het eerste onderzoek dat de effectiviteit van video-proeven evalueert in een echte onderwijscon-

text en met diverse studiejaar als criterium.

Een eerste belangrijke conclusie is dat deze studie de waarde van cognitieve voorspellers bevestigt bij de selectie van studenten voor het hoger medisch onderwijs. Wij vonden immers dat het cognitief totaal een goede voorspellende kracht had voor de studieresultaten van medische studenten. Wel dient genuanceerd te worden dat de validiteit ervan ietwat afneemt in latere jaren. We stelden dit vooral vast in curricula met minimale aandacht voor interpersoonlijke vakken. Dit was niet zo in curricula met veel aandacht voor interpersoonlijke vakken. In elk geval tonen de resultaten voor de cognitieve tests aan dat alternatieve proeven, zoals video-proeven, de traditionele cognitieve tests niet kunnen vervangen (zonder de predictieve validiteit te schaden), maar eerder aanvullen door het scala aan vaardigheden dat door traditionele tests gemeten wordt uit te breiden met niet-cognitieve vaardigheden.

Ten tweede vonden wij dat niet-cognitieve voorspellers (bijvoorbeeld een videoproef over interpersoonlijke vaardigheden) hun plaats verdienen bij de selectie van studenten voor het medisch hoger onderwijs. Een interessante conclusie hierbij is dat de voorspellende kracht van een op video opgenomen proef verschilt naargelang het type curriculum. De videoproef heeft voorspellende kracht bovenop de cognitieve voorspellers in de universiteiten waar de studieresultaten ook gebaseerd zijn op prestaties op interpersoonlijke vakken, maar niet in universiteiten waar enkel verschillen in prestaties op wetenschaps- en medische vakken een doorslaggevende rol spelen bij het berekenen van de studieresultaten. Verder vonden we dat de predictieve validiteit van de videoproef toeneemt naarmate in een vak interpersoonlijke vaardigheden worden gedoceed.

Een derde bevinding is dat in universiteiten waar de studieresultaten ook gebaseerd zijn op prestaties op interpersoonlijke vakken, de validiteit van de videoproef toeneemt naargelang studenten de medische opleiding verder doorlopen. Dit is consistent met studies die aantonen dat niet-cognitieve voorspellers aan belang winnen als de criteriumgegevens later verzameld worden (Goldstein, Zedeck, & Goldstein, 2002; Jansen & Stoop,

2001; Rothstein, Paunonen, Rush, & King, 1994). Deze resultaten in Vlaanderen kunnen een aansporing vormen voor Nederlandse universiteiten om ook alternatieve selectietechnieken te gebruiken bij de selectie voor het medisch hoger onderwijs.

Op grond van het feit dat de videoproef geen validiteit vertoonde in de groep van universiteiten waarbij de studieresultaten niet gebaseerd waren op vakken over interpersoonlijke vaardigheden, zouden sommigen ook kunnen besluiten dat deze universiteiten vanaf nu het beste prestaties op interpersoonlijke vakken ook formeel zouden moeten laten meetellen, zodat de videoproef wél een valide voorspeller zou worden. Wij gaan niet akkoord met deze redenering, omdat dit zou betekenen dat de predictor het criterium zou bepalen. Universiteiten nemen zelf strategische beslissingen over hoe zij zich van elkaar onderscheiden door bepaalde vakken meer of minder te benadrukken in hun respectievelijke curricula. De keuze van voorspellers kan deze beslissing nooit sturen.

Een laatste belangrijke bevinding is dat een videoproef - hier een op video opgenomen arts-patiëntgesprek - voorspellende kracht had, zelfs al hadden de kandidaten nooit zelf een consultatie met een patiënt uitgevoerd. Over het algemeen wordt aangenomen dat video-proeven vooral functiekennis of ervaring meten. Wat hier meespeelt, is duidelijk niet functiespecifieke kennis, maar eerder algemene kennis betreffende effectief gedrag in interpersoonlijke contacten.

Toekomstige studies moeten onderzoeken of prestaties op video-proeven gevoelig zijn voor coaching. Algemeen is het zo dat kennis van de items van de eerste vorm geen invloed mag hebben op de prestatie bij afname van een daarop volgende vorm. Als video-proeven populairder worden in toelatingsexamens, zal men proberen kandidaten een meer effectief antwoordgedrag aan te leren. Het is nog onduidelijk of coaching een gevaar vormt voor het gebruik van video-proeven bij toelatingsexamens. In tweede instantie moet toekomstig onderzoek verschillende aanbiedingsvormen met elkaar vergelijken. In deze studie werd bijvoorbeeld een op video opgenomen proef gebruikt. Het is niet duidelijk of de dure videoaanbieding een hogere voorspel-

lende validiteit heeft dan een meer goedkope, schriftelijke aanbieding.

Literatuur

- Bögels, S. M., & Kreutzkamp, R. (1990). Effecten van een training in basale gespreksvaardigheden. *Tijdschrift voor Onderwijsresearch*, 14, 201-214.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
- Goldstein, H. W., Zedeck, S., & Goldstein, I. L. (2002). Is this your final answer? *Human Performance*, 15, 123-142.
- Green, A., Peters, T. J., & Webster, J. T. (1991). An assessment of academic performance and personality. *Medical Education*, 25, 343-348.
- Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology*, 62, 429-440.
- Jansen, P. G. W., & Stoop, B. A. M. (2001). The dynamics of assessment center validity: Results of a 7-year study. *Journal of Applied Psychology*, 86, 741-753.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162-181.
- Lievens, F., & Coetsier, P. (1998). Een andere kijk op selectie van kandidaat medische studenten: Ontwikkeling van op video opgenomen simulaties. *Tijdschrift voor Hoger Onderwijs*, 13, 117-130.
- Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, 10, 245-257.
- Lievens, F., Coetsier, P., & Buyse, T. (2001). Het toelatingsexamen 'Arts en Tandarts' in Vlaanderen: Voorspellende validiteit en reacties van de studenten. *Tijdschrift voor Medisch Onderwijs*, 20, 186-195.
- Lievens, F., Coetsier, P., Janssen, P. J., & Decaestecker, C. (2001). Predictieve validiteit en genderspecificiteit van het toelatingsexamen 'Arts en Tandarts' in Vlaanderen: Een eerste peiling. *Pedagogische Studiën*, 78, 4-15.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730-740.
- Minnaert, A. (1996). *Academic performance, cognition, metacognition and motivation. Assessing freshmen characteristics on task: A validation and replication study in higher education*. Unpublished doctoral dissertation, University of Louvain, Belgium.
- Mitchell, K., Haynes, R., & Koenig, J. (1994). Assessing the validity of the updated medical college admission test. *Academic Medicine*, 69, 394-401.
- Molen, H. T. van der, & Vermeulen, W. (1995). Gedragstoetsen voor professionele gespreksvaardigheden: ontwikkeling, betrouwbaarheid en validiteit. *Tijdschrift voor Onderwijsresearch*, 20, 363-373.
- Montague, W., & Odds, F. C. (1990). Academic selection criteria and subsequent performance. *Medical Education*, 24, 44-47.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640-647.
- Powis, D. A. (1994). Selecting medical students. *Medical Education*, 28, 443-469.
- Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for restriction of range: 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.
- Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, 86, 516-530.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology.

- Journal of Applied Psychology*, 85, 112-118.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, 56, 302-318.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Smit, G. N., & Molen, H.T. van der. (1996). De constructie en evaluatie van een videotest voor de beoordeling van bekwaamheid in professionele gespreksvoering. *Tijdschrift voor Onderwijsresearch*, 21, 182-190.
- Ten Cate, T. J., 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.
- 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.
- Vey, M. A., Ones, D. S., Hezlett, S. A., Kuncel, N. R., Vannelli, J. R., Briggs, K. H., & Campbell, J.P. (2003). *Relationships among college grade indices: A meta-analysis examining temporal influences*. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- 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.
- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 25-49.
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, 52, 679-699.

Manuscript aanvaard: 4 oktober 2004

Auteurs

Filip Lievens is als hoofddocent verbonden aan de Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie van de Universiteit Gent.

Tine Buyse is als wetenschappelijk medewerker verbonden aan de Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie van de Universiteit Gent.

Correspondentieadres: F. Lievens, Vakgroep Personeelsbeleid en Arbeids- en Organisatiepsychologie, Faculteit PPW, Universiteit Gent, Henri Dunantlaan 2, Gent B-9000, België, e-mail: filip.lievens@ugent.be.

Abstract

The predictive validity of cognitive and non-cognitive predictors of the admission exam "Medical and Dental Studies" in Flanders

This study is the first large-scale examination of the validity of the Flemish Admission Exam 'Medical and Dental Studies'. The sample consisted of all students who participated in the Admission Exam between 1999-2003 and started medical studies in one of the Flemish universities. Results confirmed that cognitive predictors are valid instruments for selecting medical students. A video-based test that measured interpersonal skills was also valid in later years. Moreover, this video-based test was a valid predictor in curricula in which grade point average was also determined by courses about interpersonal skills. All of this shows that video-based tests might be an interesting and useful supplement to traditional cognitive predictors in the context of college admission.