

# Verschillen tussen de perceptie van tutoeren op de leerprestatie en de toetsprestatie van hun studenten

Drs G.A. van de Watering (gerard.vandewatering@edit.unimaas.nl) en mr S. Claessens zijn werkzaam bij de Faculteit der Rechtsgeleerdheid van de Universiteit Maastricht

*Tutores worden nogal eens geconfronteerd met studenten die ogenschijnlijk goede studenten zijn, maar die op één of andere manier niet die cijfers behalen die van hen verwacht mag worden, gezien hun bijdrage in de onderwijsbijeenkomsten. Aanleiding van dit onderzoek was een discussie tussen een aantal tutoeren, onderwijskundigen en studieadviseur over de betrouwbaarheid en validiteit van de combinatietoets. De stelling van de tutoeren was daarbij dat een toets bestaande uit alleen essayvragen beter discrimineert tussen groepen studenten dan een combinatietoets en dus eerder zal voldoen aan hun verwachtingen, in de zin dat de, in de ogen van de tutoeren, betere (competente) studenten goed presteren op de toets en de zwakkere (minder competente) studenten minder goed presteren. Doel van dit onderzoek was om na te gaan in welke mate de combinatietoets en de onderdelen van de combinatietoets (het onderdeel meerkeuzevragen en het onderdeel open vragen) onderscheid maakt tussen groepen studenten.*

*Studenten zijn aan de hand van tutorobservaties in de onderwijsgroepsbijeenkomsten van het laatste blok in de propedeuse ingedeeld in één van de volgende groepen: (1) zwakke studenten, (2) normstudenten, (3) goede studenten en (4) norm- of goede studenten met faalangst.*

*Uit de resultaten blijkt dat zowel de combinatietoets als geheel als het onderdeel meerkeuzevragen en het onderdeel open vragen op een juiste wijze onderscheid aanbrengt tussen groepen studenten. De prestaties op de onderdelen van de combinatietoets laten dezelfde trends zien, dat wil zeggen dat de zwakkere studenten slecht presteren, de normstudenten en de studenten met faalangst middelmatig en de goede studenten goed presteren. Deze resultaten zijn ook valide voor de tweede en de derde bloktoets maar niet voor de eerste bloktoets. In het geval van de eerste bloktoets presteren de faalangstige studenten beter op het onderdeel open vragen. In dit artikel wordt een verklaring voor dit verschijnsel gegeven.*

## INLEIDING

---

Studenten behalen niet altijd die resultaten op toetsen die je als onderwijsgevende mag verwachten op basis van de bijdragen van studenten tijdens de contactgebonden uren. Tutoren en studieadviseurs worden nogal eens geconfronteerd met studenten die ogenschijnlijk goede academici kunnen worden maar die op een of andere manier gehinderd worden tijdens de studievoortgang. Het is niet altijd duidelijk waardoor verschillen tussen de verwachtingen van tutoren en de uiteindelijke toetsprestatie ontstaan. Diverse onderzoeken hebben populaties studenten onderzocht en eventueel in groepen ingedeeld om verschillen in prestaties aan te tonen en te zoeken naar achterliggende factoren die verschillen in prestaties verklaren.

Deze achterliggende factoren zijn divers van aard. Over het algemeen zijn verschillen in de cognitieve capaciteit tussen studenten een belangrijke voorspeller voor verschillen in prestaties tussen studenten. Ook worden de resultaten behaald op de vooropleiding, over het algemeen gezien als een voorspellende factor voor het succes op de vervolgstudie. Er zijn echter interveniërende- en moderatorvariabelen die deze relatie kunnen beïnvloeden, zoals de sociale status van een student (Van Laar en Sidanius, 2001), de studentcultuur waaraan een student zich conformeert (Woods, 1979), het feit of een student bewust gekozen heeft voor de faculteit of studie, de academische instelling van een student (Yesilcay & Akman, 1996), de wijze waarop een student zijn studie aanpakt (Trigwell & Prosser, 1991; Vermunt, 1992), de wijze waarop een student gemotiveerd is (Biggs, 1993; Conti, 2000), de effectiviteit van de inzet van een student (Admiraal, Van den Heuvel & Gritter, 1998; Dierick & Van de Watering, 2001) en de mate waarin de student faalangst ondervindt tijdens de studie en de toets (Boekaerts, 1988; Pekrun, 1988). Ook de toetsvorm, de toetsinhoud en de wijze waarop een student de toets percipieert kan direct of indirect invloed hebben op de prestatie (Thomson & Falchikov, 1998; Scouller & Prosser, 1994; Scouller, 1998; Birenbaum, 1997; Lundeberg & Fox, 1991).

### *Toetsen van kennis en vaardigheden met de combinatie-toets*

De kennis en vaardigheden van studenten worden aan het eind van ieder blok in de propedeuse onder andere door middel van combinatie-toetsen getoetst. De combinatie-toets bestaat uit een onderdeel meerkeuzevragen (40 vragen met vier antwoordalternatieven) en een onderdeel open vragen (een essayvraag of meerdere korte antwoordvragen die zo veel mogelijk het karakter heeft of hebben van een probleemtaak zoals behandeld in de onderwijsgroepsbijeenkomsten) en is ontstaan door de voor- en nadelen van meerkeuzetoetsen en essaytoetsen ten opzichte van elkaar af te wegen. Het voordeel van de combinatie van meerkeuzevragen en essay- of open vragen is dat enerzijds zowel kennis als alle cognitieve vaardigheden getoetst kunnen worden en anderzijds zowel in de breedte als in de diepte getoetst kan worden (Van de Watering & Dierick, 2002). Driessen, Van der Vleuten & Van Berkel (1999) concludeerden reeds aan de hand van evaluaties dat de combinatie-toets goed aansluit op het probleemgestuurd onderwijs (PGO) zoals dat op de rechtenfaculteit van de Universiteit Maastricht vormgegeven is.

Een onderzoek van Dierick (2002) naar het studiegedrag van studenten ter voorbereiding op een combinatietoets toonde aan dat in de perceptie van studenten de vragen in de combinatietoets voornamelijk gericht zijn op inzicht en het kunnen toepassen van kennis en zeker niet op het reproduceren van kennis. Toch vertoonden niet alle studenten die studieactiviteiten die je mag verwachten bij deze perceptie, namelijk een grondige manier van studeren. Goede toetsresultaten werden volgens het onderzoek met name voorspeld door een grondige manier van studeren en het aanwezig zijn tijdens de onderwijsgroepsbijeenkomsten. Faalangst bleek een negatieve invloed te hebben op de prestatie; met name op het onderdeel meerkeuzevragen van de combinatietoets. De invloed van faalangst op de prestatie van verschillende toetsvormen geeft echter geen eenduidig beeld. Birenbaum en Feldman (1998) komen dan ook tot de conclusie dat de affectie voor een toetsvorm mede van invloed is op de prestatie op een toets: door een student zelf een toetsvorm te laten kiezen zal de faalangst afnemen, de motivatie toenemen en de prestatie verhoogd worden.

Uit een enquête met een kleine steekproef onder tweedejaars studenten ( $N = 52$ ) op de rechtenfaculteit van de Universiteit Maastricht bleek dat het afnemen van een combinatietoets zeer verantwoord gevonden werd in het kader van een rechtenstudie (Ciefacpol, 2000). Het merendeel van de studenten (69,2 %) gaf dan ook de voorkeur aan een combinatietoets boven een toets met enkel open vragen of een toets met enkel meerkeuzevragen. Bij deze enquête werd door studenten ook aangegeven dat de methode van toetsing van invloed is op de manier waarop gestudeerd wordt en dat het beantwoorden van open vragen gemiddeld genomen niet moeilijker is dan het beantwoorden van meerkeuzevragen.

Uit het bovenstaande blijkt dat de genoemde variabelen cognitieve capaciteit, sociale status, studentencultuur, studiekeuze, academische instelling, studieaanpak c.q. studieactiviteiten, motivatie, inzet, faalangst, perceptie op toetsen en aanwezigheid een aanzienlijke rol spelen bij de prestatie op een toets in het algemeen en een combinatietoets in het bijzonder.

#### *Doel van het onderzoek*

Doel van dit onderzoek was om na te gaan in welke mate de combinatietoets en de onderdelen van de combinatietoets (het onderdeel meerkeuzevragen en het onderdeel open vragen) onderscheid maakt tussen groepen studenten. Aanleiding van dit onderzoek was een discussie tussen een aantal tutores, onderwijskundigen en studieadviseur over de betrouwbaarheid en validiteit van de combinatietoets. De stelling van de tutores was daarbij dat een toets bestaande uit alleen essayvragen beter discrimineert tussen groepen studenten dan een combinatietoets en dus eerder zal voldoen aan hun verwachtingen, in de zin dat de, in de ogen van de tutores, betere (competente) studenten goed presteren op de toets en de zwakkere (minder competente) studenten minder goed presteren. Bovendien leggen, volgens de studieadviseur, veel studenten de schuld van het (onverwacht) slecht presteren op de combinatietoets bij het onderdeel meerkeuzevragen.

Om dit onderscheid te onderzoeken, werden studenten door middel van observaties van tutores tijdens de onderwijsgroepsbijeenkomsten aan de hand van studentprofiel-

len in groepen ingedeeld. Een studentprofiel is een kenschets van een student en bevat elementen van de eerder genoemde variabelen die de toetsprestatie kunnen beïnvloeden. Bij dit onderzoek werd zowel gekeken naar de resultaten van studenten op de gehele combinatie-toets als naar de resultaten op de onderdelen van de combinatie-toets. De eerste vraagstelling van het onderzoek richtte zich op de resultaten op de gehele combinatie-toets en luidde derhalve: maakt de combinatie-toets onderscheid tussen de door tutoren ingedeelde groepen studenten op basis van de studentprofielen? Een positief antwoord op deze vraag bevestigt tevens het vermogen van tutoren de prestatie van studenten te voorspellen aan de hand van hun indeling van studenten in studentprofielen.

Omdat de prestatie op de combinatie-toets de resultante is van de prestatie op de afzonderlijke onderdelen van de combinatie-toets, is het interessant om de prestaties van de groepen studenten op die afzonderlijke onderdelen te onderzoeken. Naar aanleiding van de door de studieadviseur en tutoren uitgesproken vermoeden was het de verwachting dat het onderdeel meerkeuzevragen minder onderscheidt maakt tussen groepen studenten dan het onderdeel open vragen en dat bepaalde studenten op het onderdeel open vragen beter presteren dan op onderdeel meerkeuzevragen. De tweede en derde vraagstelling luidden respectievelijk: maken de onderdelen van de combinatie-toets onderscheid tussen groepen studenten op basis van de door de tutoren geobserveerde karakteristieken en: is er per onderscheiden groep studenten verschil in prestatie tussen de onderdelen van de combinatie-toets?

Het onderzoek heeft in het laatste blok van de propedeuse plaatsgevonden: de tutoren hebben in dit vierde blok op basis van de studentprofielen de studenten uit hun onderwijsgroepen geïdentificeerd. Dit roept de vraag op of de gevonden verschillen en overeenkomsten tussen groepen studenten en de prestatie op (de onderdelen van) de combinatie-toets in het vierde blok ook gelden voor voorafgaande blokken. Het kan zijn dat de resultaten in blok vier blokspecifiek waren omdat studenten een geheel ander thema in die blokken bestudeerden in een onderwijsgroep met een geheel andere samenstelling die begeleid werd door een andere tutor. En bovendien is de aard van de toets niet helemaal vergelijkbaar. De toetsen kunnen gelet op het karakter van het blok verschillen in moeilijkheid en in de verhouding kennis (reproductie en inzicht) en vaardigheden (toepassen). De vierde vraagstelling van het onderzoek was afgeleid van de vraag of de gevonden overeenkomsten of verschillen blokspecifiek dan wel blokonafhankelijk zouden zijn en luidde derhalve: zijn de observaties van de onderscheiden groepen valide voor voorafgaande toetsen? Een positief antwoord op deze vraag betekent dat aan de hand van studentprofielen tutoren in een vroegtijdig stadium kunnen voorspellen welke studenten problemen zullen krijgen met de studievoortgang.

## METHODE

---

### *Subjecten*

Twee tutoren hebben aan de hand van de studentprofielen de studenten aan het eind van het blok 4 ingedeeld in groepen. Dit blok was het laatste blok van de propedeuse. Deze tutoren begeleidden in totaal 107 studenten, verdeeld over 9 onderwijsgroepen. De populatie bestond uit 56 vrouwelijke en 51 mannelijke studenten.

*Onafhankelijke variabelen*

Tutores kunnen een goed beeld krijgen van verschillen tussen studenten in de onderwijsgroepen door ze te observeren tijdens groepsdiscussies en bespreking van de taken, door de voorbereiding op de onderwijsgroepsbijeenkomsten in te zien en door studenten tijdens of na afloop van de onderwijsbijeenkomsten te vragen naar de studievoortgang. Het al dan niet goed kunnen volgen van moeilijke leerstukken is een aanwijzing voor de cognitieve capaciteit. De mate waarin studenten tijdens de (voor- of na-) bespreking van de taken relevante informatie kunnen aanhalen, de rode draad kunnen aangeven, zinvolle opmerkingen kunnen plaatsen, juiste leerdoelen kunnen formuleren, juiste verwijzingen naar (andere) bronnen kunnen geven en geïnteresseerd zijn in oplossingen of oplossingsstrategieën van andere studenten, is een indicatie van de studiehouding en studieaanpak. Studenten geven blijk van inzet en motivatie omdat ze vaak of minder vaak, goed of slecht voorbereid zijn: ze hebben de stof in meer of mindere mate doorgelezen, de geformuleerde leerdoelen uitgewerkt en extra bronnen geraadpleegd. Het niet of nauwelijks initiatief nemen bij het formuleren van leerdoelen en het zich op de vlakte houden tijdens de groepsdiscussie kunnen aanwijzingen zijn dat studenten faalangst hebben.

Aan de hand van variaties in bovengenoemde punten, die voortkomen uit de eerder genoemde variabelen die een rol spelen bij de prestatie op een toets, zijn samen met de tutores de studentprofielen geschetst waarmee groepen studenten, volgens die tutores, van elkaar onderscheiden konden worden. Vervolgens zijn de studenten door tutorobservaties in de onderwijsgroepsbijeenkomsten in de onderstaande studentprofielen ingedeeld:

Type I student: dit type student is niet of niet goed voorbereid op de onderwijsgroepsbijeenkomsten. Hij neemt weinig deel aan de discussie en zijn of haar opmerkingen tijdens de bespreking van de taken zijn weinig zinvol en bevatten weinig relevante informatie. Dit type student heeft slechts oog voor datgene wat bevraagd kan worden in de toets. De student is met name aanwezig tijdens de bijeenkomsten omwille het sociale aspect (contact met andere studenten). Deze student wordt door tutores gekenmerkt als een student met een gebrek aan (cognitieve) capaciteit en/of gebrek aan motivatie.

Type II student: dit type student kan worden getypeerd als de normstudent. Hij bereidt zich meestal en in voldoende mate voor op de onderwijsgroep. De student heeft extra uitleg nodig en voor de tutor is het voorspelbaar over welke leerstukken, namelijk de iets moeilijkere, deze student (extra) uitleg nodig heeft. De student heeft voldoende vertrouwen in zichzelf maar heeft moeite met het omschakelen van het ene leerstuk naar het andere. Hij legt zichzelf geen druk op bij het bestuderen van de leerstof.

Type III student: dit type student is cognitief zeer sterk. Hij weet uit de leerstof de rode draad te halen, kan hoofd- van bijzaken scheiden en controleert bij zichzelf of hij de stof begrijpt. Hij zoekt in andere bronnen naar verdieping of verbreding van de leerstof. Hij voelt zich erg thuis in de onderwijsgroep en neemt de verantwoordelijkheid op zich om de discussie tijdens de onderwijsgroepsbijeenkomsten goed te laten verlopen. Hij vindt het plezierig andere studenten uitleg te geven over de leerstof of om informatie uit

andere bronnen met elkaar te delen. Hij gaat zeer relaxed om met het bestuderen van de leerstof. Deze student wordt door tutores getypeerd als een adaptieve, flexibele en cognitief sterke student die bewust gekozen heeft voor deze universiteit en een PGO-leeromgeving.

In de onderwijsgroepen bevinden zich studenten die onder de type II of type III studenten zouden kunnen vallen qua voorbereiding op de onderwijsgroepsbijeenkomsten, maar wiens bijdrage aan de discussie en de bespreking van de taken beïnvloed worden door onzekerheid en krampachtigheid. Deze groep studenten kan als volgt getypeerd worden:

Type IV student: dit type student bereidt zich voor op de onderwijsgroepsbijeenkomsten, heeft de leerdoelen uitgewerkt en uitgeschreven. Tijdens de discussie en de bespreking van de taken houdt deze student zich echter op de vlakte. Hij kan echter, indien daartoe aangespoord, de kern van het probleem aanhalen en met zinvolle opmerkingen komen die bijdragen aan de discussie in de groep of aan de oplossing van de taak.

Deze studenten zijn ook zeer geïnteresseerd in wat andere studenten hebben gevonden. Zij verzamelen alle informatie die tijdens het onderwijs ter sprake komt, ook als deze informatie niet direct met de kern van de oplossing van de taak te maken heeft. Dit type student heeft het gevoel alles te moeten lezen en is daardoor veel tijd kwijt zonder dat hij vertrouwen in zichzelf krijgt en de studiestof begrijpt. Hij kan moeilijk hoofd- en bijzaken onderscheiden. Deze groep studenten worden door de tutores getypeerd als norm- of goede studenten maar met de eigenschappen faalangst, weinig flexibel en weinig adaptief.

Op basis van de tutorobservaties zijn de studenten als volgt verdeeld: type I ( $n = 19$ ), type II ( $n = 38$ ), type III ( $n = 24$ ) en type IV ( $n = 26$ ).

#### *Afhankelijke variabelen*

Voor de toets van het laatste blok van de propedeuse (periode 4) konden de studenten in beginsel in totaal 50 punten behalen: voor het onderdeel meerkeuzevragen 40 punten en voor het onderdeel open vragen 10 punten. Het onderdeel open vragen bestond in dit geval uit twee korte antwoordvragen die opgesplitst waren in totaal 5 subvragen. Naar aanleiding van de toetsanalyse na afname van de toets werd besloten één meerkeuzevraag uit de toets te verwijderen. De betrouwbaarheid van de gehele toets is acceptabel te noemen (betrouwbaarheidscoëfficiënt  $\alpha = 0,71$ ). De cesuur lag op de score 27,5 bij een maximale score van 49. Bij de vaststelling van de cesuur werd een procedure gehanteerd waarbij de grens tussen voldoende en onvoldoende berekend werd aan de hand van een absolute cesuur (60 %) met een relatief referentiepunt (95<sup>e</sup> percentiel).

Van de 107 studenten hebben 100 studenten aan deze bloktoets in periode 4 deelgenomen. In totaal hebben 458 studenten van de 501 ingeschreven studenten aan de bloktoets deelgenomen. Uit een  $t$ -toets voor een onafhankelijke steekproef blijkt dat de

100 studenten uit de steekproef gemiddeld dezelfde score behalen op de combinatie-toets als de overige 358 toetsdeelnemers.

### *Procedures*

De studenten konden zichzelf inschrijven voor een beschikbare onderwijsgroep op een bepaald tijdstip. De onderwijsgroepen werden vervolgens verdeeld over de beschikbare tutores. De studenten volgden het onderwijs in de onderwijsgroepen en legden in de 9<sup>e</sup> week na aanvang van het blok een toets af. Studenten die meer dan twee keer afwezig waren, werden niet meer toegelaten in de onderwijsgroep en verwezen naar een zogenaamde restgroep. De twee tutores hebben in de laatste week van de blokperiode hun studenten in de vier beschreven groepen ingedeeld aan de hand van hun observaties in de onderwijsgroepen. Dit indien studenten minimaal 5 van de 16 onderwijsgroepsbijeenkomsten hadden bijgewoond. Het totale onderwijs van het desbetreffende blok bestond behalve uit 16 onderwijsbijeenkomsten ook uit 8 practicumbijeenkomsten en 8 ondersteunende colleges. Naast de bloktoets werd ook een practicumtoets afgenomen.

### *Data-analyse*

Om verschillen in prestatie op de combinatie-toets en de onderdelen van de combinatie-toets tussen groepen studenten zichtbaar te maken is, naast descriptieve statistiek, gebruikgemaakt van deviatiecontrasten via de ANOVA-procedure (variantieanalyse).

## RESULTATEN

In het onderstaande beschrijven we de resultaten op basis van de diverse, eerder geformuleerde, onderzoeksvragen.

*Maakt de combinatie-toets onderscheid tussen de door tutores ingedeelde groepen studenten op basis van de studentprofielen?*

Om na te gaan of de combinatie-toets onderscheid maakt tussen groepen studenten op basis van de door tutores onderscheiden studentprofielen is per groep de gemiddelde score op de combinatie-toets berekend. De studenten presteerden als volgt op de toets:

Tabel 1 Totaalscore toets 4

studenttype	Mean	Std. Deviation	N
type I	21,31**	3,66	16
type II	26,71	6,69	36
type III	32,89**	4,46	23
type IV	29,96	5,77	25
Total	27,33	6,60	100
** $p < .01$			

Tabel 1 geeft aan dat de type I student significant lager scoort dan het gemiddelde van de overige groepen ( $t(30.54) = -6.82, p = .00$ ). Het aantal onvoldoendes in deze groep is ook het hoogst: 93,8 % van de studenten behaalt een onvoldoende. De type III student scoort duidelijk boven het gemiddelde van de overige groepen ( $t(42.98) = 7.08, p = .00$ ). Het aantal voldoende in deze groep is hier het hoogst: 87,0 %. De type II- en de type IV studenten scoren niet significant hoger of lager dan het gemiddelde van de overige groepen. Van de type II studenten behaalt 44,4 % een voldoende, van de type IV student behaalt 48,0 % een voldoende.

De studenten die als slecht gemotiveerd of die als studenten met te weinig capaciteit door de tutores aangemerkt werden (de type I studenten), presteerden onder het gemiddelde op de toets. De studenten die aangemerkt werden als normstudenten (type II studenten) scoren gemiddeld op de combinatie-toets en de studenten die door de tutores aangemerkt werden als goede studenten (de type III studenten), presteerden boven het gemiddelde. De studenten die aangemerkt werden als norm- of goede studenten maar met kenmerken van faalangst en onjuiste studievaardigheden (de type IV studenten) presteerden op hetzelfde niveau als de normstudenten.

*Maken de onderdelen van de combinatie-toets voldoende onderscheid tussen groepen studenten op basis van de door de tutores geobserveerde karakteristieken?*

Om na te gaan of de onderscheiden studenten verschillend op de onderdelen van de combinatie-toets presteren, zijn de gemiddelde scores op beide onderdelen van de combinatie-toets berekend.

De studenten scoorden op de twee onderdelen van de combinatie-toets als volgt:

Tabel 2 Score op de onderdelen van toets 4

studenttype	meerkeuzevragen			open vragen		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
type I	19,19**	3,29	16	2,12**	1,12	16
type II	22,33	5,40	36	4,38	2,47	36
type III	27,09**	3,52	23	5,80**	1,70	23
type IV	22,32	4,15	25	4,64	2,42	25
Total	22,92	5,04	100	4,41	2,39	100

\*\*  $p < .01$

Uit Tabel 2 blijkt dat de type I student significant lager scoort op het onderdeel meerkeuzevragen en het onderdeel open vraag dan het gemiddelde van de overige groepen (respectievelijk  $t(96) = -3.92, p = .00$  en  $t(41.24) = -7.61, p = .00$ ). De type III student scoort duidelijk boven het gemiddelde van de overige groepen scoort (respectievelijk  $t(96) = 5.47, p = .00$  en  $t(42,23) = 4.95, p = .00$ ). De type II student (normstudent) en de type IV studenten (faalangstige student) scoren niet significant hoger of lager op de onderdelen van de combinatie-toets dan het gemiddelde van de overige groepen.



Is er per onderscheiden groep studenten verschil in prestatie tussen de onderdelen van de combinatietoets?

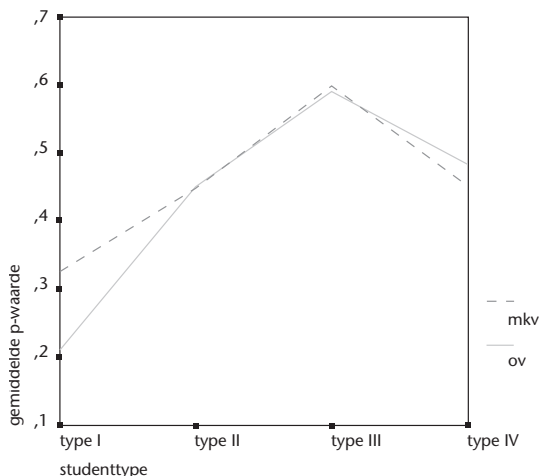
Om deze onderzoeksvraag te kunnen beantwoorden, is voor het onderdeel meerkeuzevragen en het onderdeel open vragen de gemiddelde moeilijkheidsgraad ( $p_{\text{gem.}}$ ) berekend. De gemiddelde moeilijkheidsgraad van het onderdeel meerkeuzevragen is hierbij gecorrigeerd voor het 'gedwongen raden' scoringsstelsel om een vergelijk tussen beide onderdelen beter mogelijk te maken ( $p'_{\text{gem.}}$ ).

Tabel 3 Verschil in moeilijkheidsgraad op de onderdelen van toets 4

studenttype	$p'$ gemiddeld meerkeuzevragen		p gemiddeld open vragen		p verschil		
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	N
type I	0,32	0,11	0,21	0,11	0,11**	0,14	16
type II	0,43	0,18	0,44	0,25	-0,01	0,25	36
type III	0,59	0,12	0,58	0,17	0,01	0,17	23
type IV	0,43	0,14	0,46	0,24	-0,03	0,21	25
Total	0,45	0,17	0,44	0,24	0,01	0,21	100

\*\*  $p < .01$

Uit Tabel 3 blijkt dat de moeilijkheid van beide onderdelen van de combinatietoets vergelijkbaar zijn. Om te bepalen of de onderscheiden groepen studenten een verschil in moeilijkheid ervaren op de onderdelen van de combinatietoets wordt eerst het verschil in moeilijkheidsgraad tussen die onderdelen berekend ( $p'$  meerkeuzevragen - p open vragen) en getoetst met behulp van deviatiecontrasten in de ANOVA-procedure. Hieruit volgt dat alleen de type I studenten significant verschillend presteren op de twee onderdelen van de combinatietoets: op de meerkeuzevragen presteren ze beter dan op de open vragen ( $t(28.70) = -2.82$ ,  $p = .01$ ). Figuur 1 geeft het verschil in moeilijkheidsgraad per studenttype weer.



Figuur 1 De gemiddelde p-waarde per studenttype uiteengezet voor het onderdeel meerkeuzevragen (mkv) en het onderdeel open vragen (ov).

*Zijn de observaties van de onderscheiden groepen valide voor voorafgaande toetsen?*

Om na te gaan of de gevonden resultaten blokspecifiek of blokonafhankelijk zijn, zijn de resultaten van de studenten uit de steekproef op de combinatietoets als geheel en op de onderdelen van de combinatietoets per blokperiode uiteengezet en met elkaar vergeleken. Daartoe kunnen de bovenstaande onderzoeksvragen herhaald worden op de voorafgaande blokttoetsen.

De reguliere blokttoetsen van het eerste, tweede en derde blok in de propedeutische fase hebben hetzelfde karakter als de eerder beschreven blokttoets van de vierde periode. De studenten uit de steekproef presteerden als volgt op de voorafgaande toetsen (zie tabel 4):

Tabel 4 Gemiddelde scores van de combinatietoets van de voorafgaande periodes

studenttype	toets 1			toets 2			toets 3		
	Mean	Std. Dev	N	Mean	Std. Dev	N	Mean	Std. Dev	N
type I	28,68**	3,90	17	21,21**	4,87	17	22,27**	4,22	15
type II	32,10	5,95	31	14,57	4,79	34	24,52	4,25	31
type III	33,98*	6,04	21	29,33**	5,95	23	29,30**	3,93	22
type IV	31,85	4,70	23	23,98	4,91	25	24,19	4,32	21
Total	31,83	5,54	92	24,95	5,72	99	25,24	4,79	89

\*\* p < .01

\* p < .05

Uit tabel 4 blijkt dat de type I student op alle voorafgaande combinatietoetsen als geheel significant lager scoort dan het gemiddelde van de overige groepen. De type III student scoort op alle voorafgaande combinatietoetsen significant hoger dan het gemiddelde van de overige groepen. De type II student en de type IV student scoren niet significant hoger of lager dan het gemiddelde van de overige groepen.

Om na te gaan of de onderscheiden studenten verschillend op de onderdelen van de voorafgaande combinatietoetsen presteren zijn de gemiddelde scores op beide onderdelen van de combinatietoets berekend.

De studenten uit de steekproef presteerden op het onderdeel meerkeuzevragen van de voorafgaande combinatietoetsen als volgt (zie tabel 5):

Tabel 5 Gemiddelde scores op het onderdeel meerkeuzevragen van de voorafgaande periodes

studenttype	toets 1			toets 2			toets 3		
	Mean	Std. Dev	N	Mean	Std. Dev	N	Mean	Std. Dev	N
type I	25,24*	3,67	17	18,65**	3,95	17	20,53*	3,78	15
type II	27,71	4,63	31	20,97	3,75	34	21,68	3,65	31
type III	29,24*	4,62	21	24,83**	4,89	23	25,45**	3,25	22
type IV	26,70	4,05	23	20,76	3,98	25	21,67	3,47	21
Total	27,35	4,46	92	21,41	4,56	99	22,42	3,91	89

\*\* p < .01

\* p < .05

Uit tabel 5 blijkt dat de resultaten op het onderdeel meerkeuzevragen hetzelfde beeld weergeven als de resultaten op de voorafgaande combinatie-toetsen als geheel. De gevonden resultaten op het onderdeel meerkeuzevragen blijken dus blokonafhankelijk te zijn. Het onderdeel meerkeuzevragen van de drie voorafgaande combinatie-toetsen maken op dezelfde wijze onderscheid tussen de in blok 4 door de tutoeren onderscheiden groepen studenten.

De studenten uit de steekproef presteerden op het onderdeel open vragen van de voorafgaande combinatie-toetsen als volgt (zie tabel 6):

Tabel 6 Gemiddelde scores op het onderdeel open vragen van de voorafgaande periodes

studenttype	toets 1			toets 2			toets 3		
	Mean	Std. Dev	N	Mean	Std. Dev	N	Mean	Std. Dev	N
type I	3,44**	1,70	17	2,56**	1,37	17	1,73**	1,46	15
type II	4,39	1,77	31	3,60	1,74	34	2,84	1,38	31
type III	4,74	1,96	21	4,50**	1,66	23	3,84**	1,58	22
type IV	5,15*	1,97	23	3,22	1,47	25	2,52	1,92	21
Total	4,48	1,91	92	3,54	1,70	99	2,83	1,70	89

\*\* p < .01

\* p < .05

Uit tabel 6 blijkt dat de type I student op het onderdeel open vragen van de voorafgaande combinatie-toetsen significant lager scoort dan het gemiddelde van de overige groepen. De type II student scoort niet significant hoger of lager dan het gemiddelde van de overige groepen. De type III student scoort op toets 2 en toets 3 significant hoger dan het gemiddelde van de overige groepen, maar niet op toets 1. Op toets 1 scoort deze student niet significant hoger of lager dan het gemiddelde van de overige groepen. De type IV student daarentegen scoort, afwijkend aan het niet significante resultaat van toets 2 en 3, op toets 1 significant hoger dan het gemiddelde van de overige groepen.

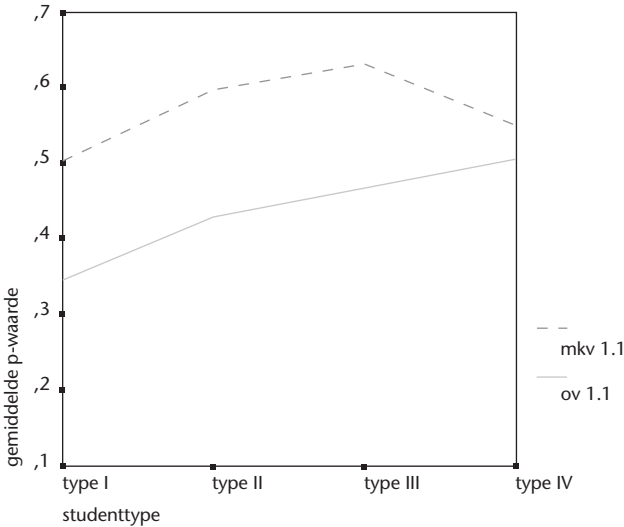
Het verschil in moeilijkheidsgraad (p' meerkeuzevragen - p open vragen) zoals deze door de onderscheiden groepen studenten wordt ervaren op de onderdelen van de combinatie-toets is in de volgende tabel aangegeven (tabel 7):

Tabel 7 Verschil in moeilijkheidsgraad tussen de onderdelen van de toets 1 tot en met 3

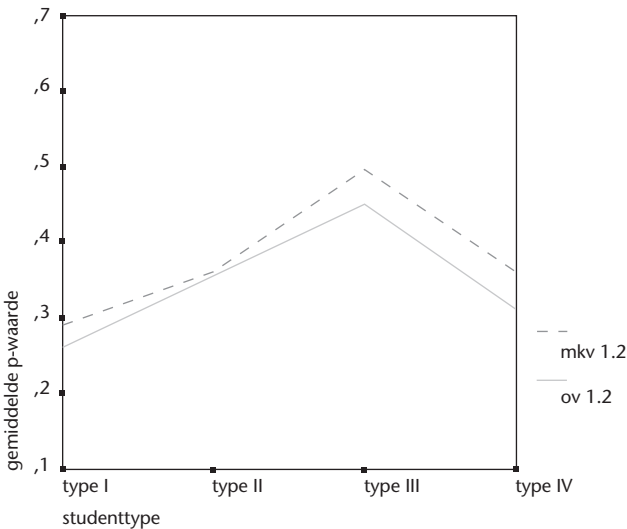
studenttype	toets 1			toets 2			toets 3		
	Mean	Std. Dev	N	Mean	Std. Dev	N	Mean	Std. Dev	N
type I	0,16	0,22	17	0,03	0,12	17	0,21	0,18	15
type II	0,15	0,14	31	0,01	0,16	34	0,14	0,16	31
type III	0,17	0,16	21	0,04	0,16	23	0,18	0,17	22
type IV	0,04*	0,23	23	0,04	0,14	25	0,17	0,20	21
Total	0,13	0,19	92	0,03	0,15	99	0,17	0,18	89

\* p < .05

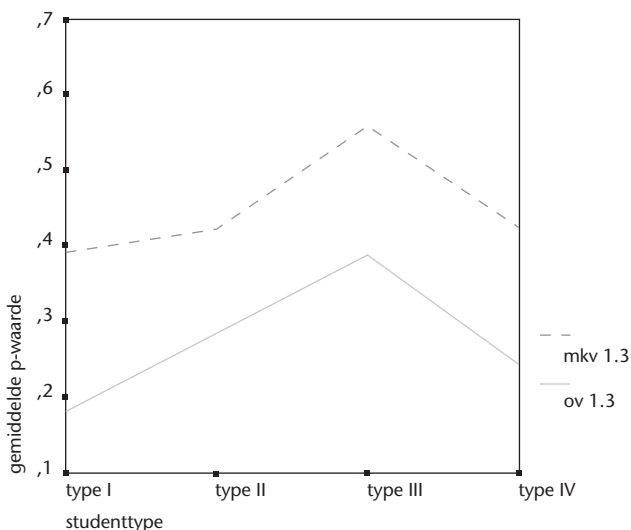
Uit tabel 7 blijkt dat alleen in het geval van toets 1 de type IV student verschillend presteert op de onderdelen van de combinatie-toets: op het onderdeel open vragen presteren deze studenten, gerelateerd aan het verschil van de overige groepen studenten tussen de twee onderdelen van de combinatie-toets, beter dan op de meerkeuzevragen. Figuur 2 tot en met 4 geven het verschil in moeilijkheidsgraad per studenttype weer voor respectievelijk toets 1 tot en met toets 3.



Figuur 2 De gemiddelde p-waarde per studenttype uiteengezet voor het onderdeel meerkeuzevragen (mkv 1.1) en het onderdeel open vragen (ov 1.1) voor toets 1.



Figuur 3 De gemiddelde p-waarde per studenttype uiteengezet voor het onderdeel meerkeuzevragen (mkv 1.2) en het onderdeel open vragen (ov 1.2) voor toets 2.



Figuur 4 De gemiddelde p-waarde per studenttype uiteengezet voor het onderdeel meerkeuzevragen (mkv 1.3) en het onderdeel open vragen (ov 1.3) voor toets 3.

## CONCLUSIE EN DISCUSSIE

De tutoeren hadden de verwachting naar aanleiding van hun observaties van het functioneren van de studenten in de onderwijsgroepen dat de combinatie-toets als geheel niet goed in staat is om studenten op een juiste wijze te discrimineren. En dat essay- of open vragen beter in staat zijn om studenten op de juiste wijze te discrimineren dan meerkeuzevragen dat kunnen. Uit dit onderzoek kan echter geconcludeerd worden dat de combinatie-toets, zoals afgenomen in het laatste blok van de propedeutische fase, voldoende onderscheid maakt tussen de door middel van de studentprofielen onderscheiden groepen studenten. Uit de resultaten van de ANOVA-procedure (tabel 2 en 3) en uit figuur 1 kan geconcludeerd worden dat de conclusies die uit de toetsprestaties op de combinatie-toets als geheel getrokken kunnen worden, ook uit de prestaties op de afzonderlijke onderdelen van de combinatie-toets getrokken worden: de studenten die als slecht gemotiveerd of als studenten met te weinig capaciteit door de tutoeren aangemerkt werden (de type I studenten), presteerden onder het gemiddelde op zowel het onderdeel meerkeuzevragen als het onderdeel open vraag. De studenten die aangemerkt werden als normstudenten (type II studenten), presteerden gemiddeld op de onderdelen van de combinatie-toets en de studenten die door de tutoeren aangemerkt werden als goede studenten (de type III studenten) presteerden op beide onderdelen boven het gemiddelde. De studenten die aangemerkt werden als norm- of goede studenten maar met kenmerken van faalangst en onjuiste studievaardigheden (de type IV studenten) presteerden op hetzelfde niveau als de normstudenten. Alleen de studenten die getypeerd werden als ongemotiveerd of met beperkte capaciteit (de type I studenten) presteren in het vierde blok op het onderdeel meerkeuze-

vragen beter dan op het onderdeel open vragen. Ook blijken tutores juist goed in staat te zijn aan de hand van hun observaties in de onderwijsgroepen de prestatie van de studenten te voorspellen door studenten in te delen in één van de beschreven studentprofielen.

De gevonden resultaten op de combinatie-toets als geheel blijken bovendien blokonafhankelijk te zijn. De drie voorafgaande combinatie-toetsen als geheel maken op dezelfde wijze onderscheid tussen de in blok 4 door de tutores onderscheiden groepen studenten (zie tabel 4). Enkel op het onderdeel open vragen blijken de gevonden resultaten niet geheel blokonafhankelijk te zijn (zie tabel 5 tot en met 7 en figuur 2 tot en met 4). Toets 1 maakt op een andere wijze onderscheid tussen de in periode 4 door de tutores gepercipieerde typen studenten. Op toets 1 presteren de type IV studenten, de studenten die door de tutores als norm- of goede studenten zijn aangemerkt met kenmerken van faalangst en onjuiste studievaardigheden relatief beter op het onderdeel open vragen dan op het onderdeel meerkeuzevragen. De type III studenten daarentegen, de studenten die door de tutores als goede studenten zijn aangemerkt, kunnen zich op deze toets op het onderdeel open vragen nog niet echt onderscheiden ten opzichte van de rest van de studenten.

De discussie is hier op welke wijze de perceptie van studenten de prestatie op de combinatie-toets beïnvloedt. De prestatie van de studenten op de toetsen is vrij stabiel, maar de eerste toets is duidelijk bepalend voor met name de type IV-student. Een mogelijke verklaring ligt hier in de zogenaamde pre-assessmenteffecten (Gielen, Dochy & Dierick, 2003). Met deze term wordt de aanwezige invloed van de toets op het leergedrag van studenten aangeduid, in het bijzonder de wijze waarop studenten zich voorbereiden op de toets. Volgens Boud (1990) focussen studenten zich op die onderwerpen en verwerkingsniveaus die aan bod komen in de toets en die volgens hen punten opleveren. Er bestaat daarbij een discrepantie tussen wat er daadwerkelijk op de toets komt en wat naar verwachting van de studenten op de toets komt (Broekkamp, 2002).

In geval van de type IV-studenten is er sprake van een verschuiving in de perceptie op de toets die op zijn beurt invloed heeft op de toetsresultaten. Na de ervaring met de eerste combinatie-toets, waarin het onderdeel open vragen relatief beter wordt gemaakt dan het onderdeel meerkeuzevragen, wordt de waardering voor de onderdelen van de toets bijgesteld. Ondanks het feit dat het onderdeel open vragen door deze studenten goed gemaakt wordt, valt het totaalresultaat waarschijnlijk zo tegen dat ze zich vooral gaan richten op dat andere onderdeel van de toets: de meerkeuzevragen. Dit werpt een nieuw licht op de conclusie van Birenbaum (1997) dat faalangst een overheersende factor is die ervoor zorgt dat studenten meer neigen naar een voorkeur voor meerkeuzevragen. Dit effect is beter te verklaren met de eerder genoemde observaties van Boud: studenten zijn bereid hun studieaanpak te verleggen indien ze er van overtuigd zijn dat daardoor hun kans op slagen wordt verhoogd. Het lijkt erop dat de consequentie voor de studenten met faalangst (type IV) is dat ze hun aandacht verleggen van het studeren gericht op open vragen naar studeren gericht op meerkeuzevragen en daarbij een meer oppervlakkige manier (met name memoriseren) van studeren gaan hanteren en onjuiste studievaardigheden gaan toepassen (geen onderscheid maken tussen hoofd-

en bijzaken enzovoort). Dit zou een verklaring kunnen zijn dat het resultaat op het onderdeel open vragen in de volgende toetsen minder wordt. In dit geval kan men spreken van een negatieve invloed van het pre-assessmenteffect op het studiegedrag. Bij de type III studenten zou men juist kunnen spreken over een positieve invloed van het pre-assessmenteffect omdat deze studenten na de eerste toets beter gaan presteren op het onderdeel open vragen.

Verder vervolgonderzoek zou zich kunnen richten op de wijze hoe de te onderscheiden groepen studenten de combinatie-toets percipiëren en op welke wijze de perceptie op de toets beïnvloed kan worden, bijvoorbeeld door het gewicht van de meerkeuzevragen en de open vragen te veranderen. Daarbij is het noodzakelijk om in een zo vroeg mogelijk stadium (in periode 2) de type IV-studenten op te sporen en te helpen bij de voorbereiding op de toets om te voorkomen dat ogenschijnlijk goede studenten gehinderd worden in hun studievoortgang.

## LITERATUUR

---

- Admiraal, W., Heuvel, T. van den & Gritter, M. (1998) Studieprestaties en interesse voor de studie Rechten, *Tijdschrift voor Hoger Onderwijs*, 16, 131-144.
- Biggs, J.B. (1993) What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology*, 63, 1-17.
- Birenbaum, M. (1997) Assessment preferences and their relationship to learning strategies and orientations. *Higher education*, 33, 71-84.
- Birenbaum, M. & Feldman, R.A. (1998). Relationships between learning patterns and attitudes towards two assessment formats, *Educational Research*, 40, 90-98.
- Boekaerts, M. (1988) Motivated learning: bias in appraisals. *Studies in Educational Evaluation*, 12, 267-280.
- Boud, D. (1990) Assessment and the promotion of academic values. *Studies in Higher Education*, 15, 101-111.
- Broekkamp, H. (2002) *Task demands and test expectations*. (Academisch proefschrift). Amsterdam: Universiteit van Amsterdam.
- Ciefacpol. (2000) *Evaluatie Enquête-resultaten toetsmethode*. Maastricht: Commissie Facultaire Politiek der Juridische Faculteitsvereniging Ouranos.
- Conti, R. (2000) College goals: Do self-determined and carefully considered goals predict intrinsic motivation, academic performance, and adjustment during the first semester? *Social Psychology of Education*, 4, 189-2111.
- Dierick, S. (2002) *Onderzoek naar het leergedrag van studenten ter voorbereiding van de bloktoets 1.3 en hun percepties aangaande bepaalde toetskwaliteiten*. Maastricht: EDIT.
- Dierick, S. & Watering, G. van de (2001) *Toetsanalyse: Reguliere toets 2.3 "Staats - en bestuursrecht"*. Maastricht: EDIT.
- Driessen, E., Vleuten, C. van der & Berkel, H. van (1999) Beyond the multiple-choice v. essay questions controversy. *The Law Teacher*, 33, 159-171.
- Gielen, S., Dochy, F. & Dierick, S. (2003) Evaluating the consequential validity of new modes of assessment: the influence of assessment on learning, including pre, post and

- true assessment effects. In: M. Segers, F. Dochy & E. Cascallar (eds.), *Optimizing new modes of assessment: in search of qualities and standards*. Boston: Kluwer academic, 37-54.
- Lundeberg, M.A. & Fox, P.W. (1991) Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, 61, 94-106.
- Pekrun, R. (1988) Anxiety and motivation in achievement setting: towards a system-theoretical approach. *Studies in Educational Evaluation*, 12, 307-323.
- Scouller, K. (1998) The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher education*, 35, 453-472
- Scouller, K.M. & Prosser, M. (1994) Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education*, 19, 267-269.
- Thomson, K. & Falchikov, N. (1998) "Full on Until the Sun Comes Out": the effects of assessment on student approaches to studying. *Assessment & Evaluation in Higher Education*, 23, 379-390.
- Trigwell, K. & Prosser, M. (1991) Relating learning approaches, perceptions of context and learning outcomes, *Higher Education*, 22, 252-266.
- Laar, C. van. & Sidanius, J. (2001) Social status and the academic achievement gap: A social domain perspective, *Social Psychology of Education*, 4, 235-258.
- Vermunt, J.D.H.M. (1992) *Leerstijlen en sturen van leerprocessen in het hoger onderwijs*. Amsterdam: Swets & Zeitlinger.
- Watering, G. van de & Dierick, S. (2002) Kwaliteit van assessment: de bruikbaarheid van klassieke toetsen binnen studentgericht- en competentiegericht onderwijs. In: F. Dochy, L. Heylen & H. Van de Mosselaer (red). *Assessment in onderwijs*. Utrecht: Lemma, 61-89.
- Woods (1979), P. (1979) *The divided school*. London: Routledge and Kegan.
- Yesilcay, Y. & Akman, K.I. (1996) A statistical model for the early detection of achievers and non-achievers in a university. *Studies in Educational Evaluation*, 22, 59-77.