



Universiteit
Leiden
The Netherlands

Presteren in het VWO en het HO. Deelonderzoek van experimenten met selectie: selectie op basis van vooropleidingsgegevens

Gruijter, D.N.M. de; Yildiz, M.; t Hart, J.; Béguin, A.A.

Citation

Gruijter, D. N. M. de, Yildiz, M., T Hart, J., & Béguin, A. A. (2005). Presteren in het VWO en het HO. Deelonderzoek van experimenten met selectie: selectie op basis van vooropleidingsgegevens. Retrieved from <https://hdl.handle.net/1887/7808>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/7808>

Note: To cite this publication please use the final published version (if applicable).

Presenteren in het VWO en het HO

Deelonderzoek van experimenten met selectie: selectie op basis van vooropleidinggegevens

Rapport nr 148
Maart 2005

D.N.M. de Gruijter, ICLON
M. Yildiz, Bestuursbureau
&
J. 't Hart, ICS
Met een bijdrage van A.A. Béguin, CITO



Presteren in het VWO en het HO

*Deelonderzoek van experimenten met selectie: selectie op basis van
vooropleidinggegevens*

Rapport 148

Maart 2005

D.N.M. de Gruijter, ICLON

M. Yildiz, Bestuursbureau

&

J. 't Hart, ICS

met een bijdrage van A.A. Béguin, CITO

Inhoudsopgave

	Samenvatting en conclusies	3
1.	Inleiding	7
2.	Opzet van het onderzoek	9
2.1	De succescriteria	9
2.2.	De te onderzoeken groepen	10
2.3.	De voorspellers	10
3.	De cijfers voor het VWO nieuwe stijl	13
3.1	De profielen	13
3.2	Het cijfermateriaal	15
3.3	Het centraal schriftelijk versus het eindcijfer	16
3.4	De verplichte onderdelen en de keuzeruimte bij een deelgroep	18
4.	De cohorten	21
5.	De relatie tussen gemiddelde eindexamencijfers en het al dan niet behalen van de propedeuse na 1, resp. 2 jaar	25
5.1	Psychologie; cohorten 2002 en 2003	25
5.2	Bestuurskunde; cohorten 2001, 2002 en 2003	29
5.3	Geschiedenis; cohorten 2001, 2002 en 2003	31
5.4	Rechten; cohorten 2001, 2002 en 2003	33
5.5	Drie talenstudies	34
5.6	De verschillende cijfergemiddelden	36
6.	De relatie tussen eindexamencijfers en het aantal behaalde studiepunten	43
7.	Het vergelijkbaar maken van cijfers op heelvakken en deelvakken	53
8.	Literatuur	61
Bijlage 1:	De cijfers bij de opleiding en het behalen van de propedeuse in 1 en 2 jaar	
Bijlage 2:	Het koppelen van de examens van het deelvak en het heelvak	
	A.A. Béguin	

Samenvatting en conclusies

❖ Aard van het onderzoek

In dit onderzoek is voor enkele opleidingen van de Universiteit Leiden nagegaan hoe sterk de relatie is tussen VWO-cijfers en studiesucces. De conclusies die op basis van het onderzoek worden getrokken, betreffen uiteraard alleen de onderzochte opleidingen.

❖ De onderzochte groepen

Het onderzoek had betrekking op de opleidingen Bestuurskunde, Geschiedenis, Psychologie en Rechten, en de drie talenstudies Duits, Frans en Italiaans. Onderzocht werden de groepen studenten uit de cohorten voltijdstudenten 2001/2002, 2002/2003 en 2003/2004 van wie VWO-cijfers bekend waren.

❖ De predictoren

- De VWO-eindexamencijfers vormden het startpunt van het onderzoek. De VWO-cijfers zijn opgeslagen in het registratiesysteem van de universiteit. Bij controle van de gegevens bleek dat er bij de opslag van gegevens fouten mogelijk zijn.
- Voor een deel van de studenten bleek het mogelijk de VWO-cijfers opnieuw in te voeren. Voor deze groep studenten zijn de cijfers voor de centraal schriftelijke examens apart opgeslagen.
- Voor het onderzoek werd gebruik gemaakt van verschillende voorspellers die op de VWO-cijfers zijn gebaseerd. De eerste predictor is het (geschatte) gemiddelde cijfer dat bij de procedure voor de gewogen loting wordt gebruikt. Voor een deelgroep van studenten is het gemiddelde cijfer voor het centraal schriftelijk en het gemiddeld eindcijfer over vakken met een centraal schriftelijk berekend. Voorts werden de cijfers voor enkele afzonderlijke vakken gebruikt.
- Een specifiek deel van het onderzoek betrof het cijfer voor Wiskunde. Hiertoe werden de cijfers voor Wiskunde A1, Wiskunde A1,2, Wiskunde B1 en Wiskunde B1,2 uit de profielen van het vernieuwde VWO op een gemeenschappelijke schaal gebracht, op basis van onderzoek dat door het CITO is verricht naar de relatie tussen heelvakken en deelvakken.

❖ De criteria

- Als succes binnen de opleiding werd het al dan niet behalen van de propedeuse binnen één, respectievelijk twee jaar gehanteerd. Het al dan niet binnen twee jaar behalen van de propedeuse sluit aan bij de norm die door de universiteit van af 2005-2006 aan voltijdstudenten wordt gesteld. Het moge duidelijk zijn dat succes in de verdere studie afhankelijk is van het tijdig behalen van de propedeuse. Het al dan niet behalen van de propedeuse in één jaar is een nuttig criterium aangezien dit criterium al na één jaar beschikbaar is. Bovendien zijn studenten die in één jaar de propedeuse halen, gemiddeld beter dan studenten die de propedeuse later halen.
- Een derde criterium dat werd gehanteerd is het aantal behaalde studiepunten (ECTS) aan het eind van het eerste studiejaar. Dit criterium geldt als een maat voor studiesnelheid.
- Gegevens m.b.t. de criteria zijn op centraal niveau beschikbaar. De gegevens m.b.t. het aantal behaalde ECTS worden t.b.v. het Bindend StudieAdvies aangeleverd door de opleidingen. Uit het onderzoek bleek dat de gegevens m.b.t. het aantal ECTS niet in alle gevallen correct waren.

❖ De resultaten

- De relatie tussen VWO-cijfers en studiesucces verschilt van opleiding tot opleiding. In alle gevallen is echter duidelijk dat studenten met een relatief laag gemiddeld eindexamencijfer minder succesvol in de studie zijn.
- Bij de opleidingen Psychologie, Geschiedenis en Rechten werden bij een deelgroep van studenten extra gegevens berekend. Daarbij bleek een gemiddeld cijfer voor vakken die een centraal examendeel kennen, een goede voorspeller op te leveren. Dit gemiddelde cijfer bleek studiesucces beter te voorspellen dan een gemiddeld cijfer gebaseerd op alleen de centrale toetsen van deze vakken.
- Bij een aantal opleidingen is gekeken naar de samenhang tussen studiesucces en het cijfer voor een VWO-vak dat relevant voor de universitaire opleiding geacht wordt. Bij Duits was de voorspellende waarde van het VWO-cijfer Duits groter dan de voorspellende waarde van een gemiddeld eindexamencijfer, in termen van de correlatie met studiesucces. Het onderzoek bij Duits was gebaseerd op een klein aantal studenten. Bij de overige opleidingen bleek het beter uit te gaan van een gemiddeld VWO-cijfer als voorspeller van studiesucces. Vervolgens

kan dan wel worden gekeken of toevoeging van een cijfer voor één van de afzonderlijke VWO-vakken de voorspelling van studiesucces verbetert. Bij een aantal opleidingen is dat onderzocht. Het bleek dat de voorspelling niet significant verbeterde als ook gebruik werd gemaakt van cijfers voor een specifiek VWO-vak. Daarvoor zijn enkele redenen. Het feit dat sommige onderzochte groepen klein zijn, houdt in dat het niet gemakkelijk is statistisch significante effecten te verkrijgen. T.a.v. vakken die verwantschap vertonen met vakken uit de universitaire opleiding kan zelfselectie tot een beperking van de range van cijfers leiden, met als gevolg een neerwaarts effect op de correlatie met studiesucces.

- Het aantal in het eerste studiejaar behaalde studiepunten lijkt ook een nuttig criterium in selectie-onderzoek, een criterium dat sneller gehanteerd kan worden dan het propedeuserendement na twee jaar. De procedure die momenteel wordt gehanteerd om het aantal studiepunten vast te leggen, vertoont echter, zoals vermeld, grote tekortkomingen.

❖ Conclusies

- Het gemiddelde van de cijfers op de centrale schriftelijke examens en het gemiddelde cijfer van de vakken met een centraal schriftelijk kunnen zijn beide als voorspeller van studiesucces gehanteerd worden. Bij een keuze tussen deze twee opties speelt niet alleen de hoogte van de voorspellende waarde een rol, maar ook de vraag in hoeverre het acceptabel is om niet gestandaardiseerde schoolonderzoeken bij de selectie een rol te laten spelen. De invulling van de profielen zal in de toekomst overigens veranderen. Als de rol van de centrale examens daarbij niet verandert, mag er vanuit gegaan worden dat de conclusies die hier m.b.t. de voorspellers zijn getrokken, naar de toekomst generaliseerbaar zijn.
- Gezien de samenhang tussen VWO-cijfers en studiesucces ligt het voor de hand leerlingen met gemiddeld hoge VWO-cijfers bij selectie rechtstreeks tot de studie van hun keus toe te laten. Er zal wel bepaald moeten worden over welke cijfers gemiddeld wordt; zie hiervoor het voorgaande punt.
- Indien er geselecteerd gaat worden, heeft dat ongetwijfeld effect op de samenstelling van de groep belangstellenden. Leerlingen met een laag gemiddeld VWO-cijfer zouden eerder kunnen afzien van een poging de desbetreffende opleiding te gaan doen.

- Een geringe samenhang tussen een voor de hand liggende voorspeller zoals het cijfer voor een moderne taal en studiesucces in de desbetreffende studierichting betekent niet dat die voorspeller niet zou moeten worden gebruikt. Het kan voor de hand liggen om bij selectie voor talenstudies indien mogelijk een taleneis te stellen, teneinde het karakter van de opleiding duidelijk te maken. Een lage samenhang, zoals gevonden in het huidige onderzoek, houdt alleen in dat het rendement van een dergelijke selectie verwaarloosbaar is en dat andere instrumenten als de feitelijke selectie-instrumenten in het selectieproces moeten worden betrokken.
- Studenten met een ander dan een VWO-diploma vielen buiten dit onderzoek. Indien men op basis van cijfers zou willen selecteren, zal men bij niet VWO'ers een andere basis voor de voorspelling van studiesucces moeten kiezen.
- Voor kandidaten zonder VWO-diploma is er binnen het systeem van gewogen loting een oplossing gevonden: niet VWO'-ers worden automatisch in een lotingsklasse (lotingsklasse C) ingedeeld. Onderzoek bij Psychologie wees uit dat bij deze indeling – althans bij Psychologie – niet VWO'ers als groep te hoog worden ingedeeld, uitgaande van het studiesucces van de verschillende 'lotingsklassen'.

1. Inleiding

In maart 2004 bracht de *Taskforce Capaciteit en Kwaliteit* een advies uit aan het College van Bestuur over selectie voor en tijdens het onderwijs. In het advies wordt een pleidooi gehouden voor experimenten met selectie aan de poort bij een aantal opleidingen aan de universiteit. Op basis van deze adviezen besloot het College vervolgens tot de start van het project *Decentrale selectie*. Doel van het project is verhoging van de kwaliteit van de instroom en het bereiken van een hoger studierendement.

Er werd besloten bij een beperkt aantal opleidingen onderzoek te doen naar de mogelijkheden om goede studenten voor de opleiding te selecteren en niet geschikte kandidaten voor de desbetreffende opleiding af te wijzen (deze kandidaten kunnen er vervolgens wel voor kiezen bij een andere Leidse opleiding in te stromen).

De bij het onderzoek betrokken opleidingen zijn: Bestuurskunde, Duits, Frans, Geschiedenis, Italiaans, Psychologie, en Rechten. Welke instrumenten kunnen gebruikt worden als selectie-instrumenten? De instrumenten moeten voldoen aan eisen van validiteit, acceptabiliteit en de kosten moeten worden afgezet tegen de baten¹. Bij een selectiestudie is natuurlijk de predictieve validiteit van groot belang. Bij predictieve validiteit gaat het om de mate waarin een selectieprocedure erin slaagt potentieel succesvolle kandidaten aan te wijzen. Een bruikbare predictieve waarde van een voorspeller is echter niet voldoende voor de acceptatie van een meetinstrument als selectie-instrument. Bij acceptabiliteit van een procedure voor alle betrokkenen gaat het ook om factoren als *face validity* en de *consequential basis* van het testgebruik, d.w.z. welke positieve en negatieve effecten heeft het gebruik van een meetinstrument in een bepaalde context (Messick, 1989).

Het ligt voor de hand bij de selectieprocedures ten behoeve van het universitaire onderwijs (ook) naar VWO-eindexamencijfers te kijken. Eindexamencijfers hebben een zekere predictieve waarde: de eindexamencijfers zijn gebaseerd op veelal recente prestaties die inzet en capaciteiten vergen die vergelijkbaar zijn met die welke nodig zijn voor universitaire opleidingen (Commissie Toelating Numerus Fixusopleidingen, 1997; Scherft, J. P., & Van Hilten, W., 1980). Het is van enig

¹ De 'opbrengst' van de selectie hangt o.a. af van de base rate, het percentage geschikte kandidaten, en van de selectieratio, het percentage kandidaten dat wordt toegelaten.

belang dat de kosten van een procedure gebaseerd op eindexamencijfers, relatief laag zijn.

De keuze voor eindexamencijfers is overigens niet geheel zonder problemen. Het grootste probleem is dat een gedeelte van de instroom aan de universiteit binnen komt met een andere vooropleiding dan het VWO, bijvoorbeeld via het HBO, na een HAVO-opleiding; voor hen moet op andere gegevens worden geselecteerd. Een tweede probleem is dat de cijfers van verschillende kandidaten niet geheel vergelijkbaar zijn: leerlingen doen verschillende vakken en de eindexamencijfers berusten niet alleen op een gemeenschappelijk landelijk examen, maar voor een deel ook op een schoolexamen, ook al is het probleem minder groot dan in Vlaanderen, dat geen centraal eindexamen kent, en waar voor enkele studies een toelatingsexamen bestaat (Lievens, F., Coetsier, P., Janssen, P.J., & Decaesteker, C., 2001) en daarnaast als hulpmiddel bij de studiekeuze de 'Chrysostomos'-uitstaptoets voor het secundair onderwijs is ontwikkeld (Janssen, 2004). Het tweede probleem zal in het huidige onderzoek besproken worden.

Feitelijk worden VWO-eindexamencijfers al gebruikt voor instroomregulatie. Bij bijvoorbeeld Geneeskunde is de belangstelling groter dan de landelijke opleidingscapaciteit en wordt de instroom beperkt via het systeem van gewogen loting, waarbij voor de VWO'ers eindexamencijfers tellen. In 2004 is ook bij Psychologie in Leiden voor het eerst een instellingsfixus gehanteerd nadat in het studiejaar 2003-2004 de instroom groter was dan de opleidingscapaciteit. Voor Psychologie bleef de instroom in 2004 echter onder de maximaal toegestane instroom; wellicht is de afname mede het gevolg van een afschrikwekkend effect van de ingestelde toelatingsbeperking.

In het onderhavige onderzoek worden de mogelijkheden van een voorspelling van het studiesucces op basis van VWO-cijfers onderzocht voor de genoemde zeven opleidingen. Daarbij past de kanttekening dat de talenopleidingen die bij het onderzoek betrokken zijn, een geringe instroom hebben.

2. Opzet van het onderzoek

Voor het onderzoek moeten relevante cohorten studenten worden onderzocht. Voorts moet worden besloten op welke criteria en voorspellers de studenten moeten worden vergeleken. In de volgende paragrafen komen de keuzes aan de orde.

2.1 De succescriteria

Er zijn verschillende criteria voor succes in de studie mogelijk. Uiteindelijk gaat het om het al dan niet halen van de bachelorgraad, het liefst binnen een beperkt aantal jaren. Daarvoor is wel nodig dat de propedeuse wordt gehaald. De universiteit gaat de studievoortgang strenger aan te pakken dan via het huidige BSA met een norm van 30 ECTS. De norm wordt verhoogd naar 40 ECTS en er gaat van voltijdstudenten geëist worden dat zij de propedeuse binnen twee jaar halen. Het al dan niet behalen van de propedeuse binnen twee jaar lijkt daarom een geschikt criterium. Het criterium sluit aan bij het uitgangspunt van de studie bij Psychologie. Tevens zal worden gekeken naar het al dan niet behalen van de propedeuse binnen de nominale studieduur van de propedeuse, één jaar. Studenten die de propedeuse binnen de nominale studieduur halen, mogen daardoor tot de beter studenten gerekend worden. Het al dan niet behalen van de propedeuse in één jaar is ook in andere studies als criterium gehanteerd (Meerum Terwogt-Kouwenhoven, 1980; Scherft & Van Hilten, 1980). Roeleveld (1997) onderzocht in zijn empirische studie naar het systeem van gewogen loting de relatie tussen lotingscategorieën en het al dan niet slagen voor de propedeuse (ongelimiteerd). Hij onderzocht bovendien de relatie tussen lotingscategorieën en al dan niet slagen voor latere fases in de studie.

Er zal in het onderzoek ook een tweede criterium worden meegenomen: het aantal voor het BSA meetellende ECTS aan het eind van het eerste studiejaar, als maat voor studietempo. Dit criterium is ook in verschillende eerdere studies gehanteerd (Meerum Terwogt-Kouwenhoven, 1980; Smits, Mellenbergh, & Vorst, 2002). Roeleveld (1997) onderzocht voor de deelgroepen van geslaagden ook de studiesnelheid van de studenten in termen van het aantal maanden nodig voor het behalen van de propedeuse.

2.2. De te onderzoeken groepen

Een volgende keuze die moet worden gemaakt is die naar de cohorten die moeten worden gevolgd. Als het al dan niet halen van de propedeuse na twee jaar een succescriterium is, moeten in ieder geval cohorten worden genomen die minstens twee jaar met de studie bezig zijn geweest. Bij een start van het onderzoek na 1 september 2004 zijn dat de cohorten 2002 en eerdere cohorten. Het is van belang andere cohorten dan de cohorten 2002 in het onderzoek te betrekken: er kunnen dan namelijk uitspraken worden gedaan over de stabiliteit van de gevonden resultaten. Er zijn echter eveneens goed redenen om cohorten die lang geleden met de studie zijn begonnen, niet in het onderzoek te betrekken. Immers, het VWO is de laatste jaren veranderd: het studiehuis is ingevoerd. Ook de te onderzoeken universitaire opleiding kan het propedeuseprogramma hebben aangepast. Het ligt daarom voor de hand het onderzoek te beperken tot de meest recente cohorten. Dat zijn de cohorten 2001 en 2002, de twee cohorten waarbij gegevens na twee jaar bekend zijn, en de cohorten 2003 waarbij het propedeuserendement na één jaar en het aantal behaalde ECTS in verband met het BSA na één jaar bekend zijn.

In de studie wordt het aantal behaalde BSA-ECTS als tweede criterium gehanteerd. De BSA-cohortdefinitie wijkt af van de cohortdefinitie volgens de 1 decembertellingen. Studenten die zich voor 1 maart van het eerste studiejaar uitschrijven, worden niet meegenomen in de uiteindelijke BSA-cohortdefinitie. Wij zullen de groep van studenten die niet in de BSA-cohort zitten, als aparte deelgroep meenemen in het onderzoek naar de relatie tussen het aantal behaalde BSA-ECTS, het propedeuse rendement en VWO-cijfers.

2.3. De voorspellers

Zoals gezegd, heeft dit onderzoek betrekking op vooropleidingresultaten. Van studenten die op basis van het VWO diploma instromen, zijn bij het eindexamen behaalde cijfers bekend. Voor studenten die op andere wijze instromen, zijn dergelijke gegevens niet beschikbaar. Bij de huidige instroomselectie bij numerus fixus opleidingen wordt dat probleem ‘opgelost’ door het toewijzen van gegadigden zonder VWO diploma aan een bepaalde lotingscategorie (categorie C). In dit onderzoek kijken wij vooral naar de deelgroep van studenten met een VWO-diploma en met name naar studenten die het VWO nieuw stijl hebben gedaan en tijdens de laatste jaren van hun VWO-opleiding een bepaald profiel hebben

gevolgd. In het volgende hoofdstuk wordt nader op de eindexamencijfers voor het VWO nieuw stijl ingegaan.

3. De cijfers voor het VWO nieuwe stijl

3.1 De profielen

De diploma's van het eindexamen VWO zijn tegenwoordig gebaseerd op de resultaten van een gemeenschappelijk deel, een profieldeel en de vrije ruimte. Er zijn vier profielen, te weten:

Cultuur & Maatschappij (C&M)

Economie & Maatschappij (E&M)

Natuur & Gezondheid (N&G)

Natuur & Techniek (N&T).

In het algemeen gedeelte vallen slechts twee vakken, Nederlands en Engels, die zowel een schoolexamen als een centraal schriftelijk kennen. De overige vakken hebben alleen een schoolexamen. Het aantal vakken in het algemeen gedeelte varieert. In het algemeen gedeelte zit bijvoorbeeld het vak Frans 1. Leerlingen die in de laatste klassen doorgaan met Frans, doen Frans 1,2 en dat vak is of een profielvak (mogelijk bij *C&M*) of valt onder de vakken in de vrije ruimte.

De profielen hebben drie (*N&T*) tot vijf (*C&M*) vakken. De vakken in de profielen worden ook centraal geëxamineerd met uitzondering van CKV2 als onderdeel van CKV2,3 in het profiel *C&M*.

Voor de profielen *Economie & Maatschappij*, *Natuur & Gezondheid* en *Natuur & Techniek* legt de profielkeuze de vakken voor het profiel vast². Dat is niet het geval bij het profiel *Cultuur & Maatschappij*. Bij dit profiel is de keuze van de vakken voor het profieldeel van het eindexamen ruim. De vrije ruimte biedt veel mogelijkheden. Door een goede keuze van extra vakken zijn er leerlingen die voor het eindexamen in meer dan één profiel slagen. Zo is het gemakkelijk om naast het profiel *N&T* met de vakken Wiskunde B 1,2, Natuurkunde 1,2, Scheikunde 1,2 het profiel *N&G* te doen: daartoe moet ook het vak Biologie 1,2 worden afgerond.

Bij sommige vakken is er alleen een schoolexamen, bij andere vakken is het eindcijfer van het eindexamen het (afgeronde) gemiddelde van een centraal examen en een schoolexamen. Een deel van de vakken is al ruim voor het eindexamen

² Dat is zo bij de huidige eindexamens. De invulling van de profielen wordt in de toekomst echter gewijzigd.

afgerond. Bij enkele onderdelen worden geen cijfers gegeven, maar een waardering zoals Voldoende.

Hoe kunnen leerlingen met uiteenlopende vakkenpakketten vergeleken worden? Bij de aanmelding voor studies met een landelijke of plaatselijke numerus fixus wordt door de IB-groep een gemiddeld eindexamencijfer berekend waarbij de cijfers voor het gemeenschappelijk deel en het profieldeel meetellen, evenals het hoogste cijfer uit de vrije ruimte³. De gemiddelde cijfers worden in een aantal lotingsklassen verdeeld: leerlingen met een 8 of hoger gemiddeld stromen rechtstreeks in (lotingsklasse A), een gemiddeld cijfer lager dan 8 maar hoger dan of gelijk aan 7.5 valt in klasse B, een gemiddeld cijfer lager dan 7.5 maar hoger dan of gelijk aan 7 valt in lotingsklasse C, een gemiddelde lager dan 7 maar hoger dan of gelijk aan 6.5 valt in lotingsklasse D, een lager gemiddelde valt in lotingsklasse E. Deze berekening maakt geen onderscheid tussen prestaties bij verschillende profielen (Natuurkunde 1 versus Natuurkunde 1,2 of Latijn versus Natuurkunde 1,2), tussen prestaties voor deelvakken en heelvakken (Natuurkunde 1 versus Natuurkunde 1,2), tussen prestaties voor onderdelen die niet centraal worden getoetst en prestaties voor onderdelen die mede centraal worden getoetst (Frans 1 versus Frans 1,2), tussen prestaties voor onderdelen die in het verleden zijn afgesloten en prestaties voor onderdelen die aan het eind van 6 VWO worden afgesloten (Frans 1 versus Frans 1,2), en tussen cijfers die in verschillende jaren zijn toegekend (Frans 2003 versus Frans 2002). Het is dus heel wel mogelijk dat de voorspelling van studiesucces bij een bepaalde universitaire opleiding op basis van eindexamencijfers verbeterd kan worden vergeleken met een voorspelling gebaseerd op het gemiddelde cijfer dat de IB-groep hanteert. Er is relevant onderzoek gedaan. Smits, Mellenbergh en Vorst (2002) toonden aan dat een gemiddelde waarbij cijfers voor niet geëxamineerde vakken met statistische technieken worden ingevuld op basis van beschikbare cijfers, een betere voorspelling kan opleveren dan een gewoon cijfergemiddelde. Hun onderzoek beperkte zich echter tot de verplichte vakken Nederlands en Engels en drie overige vakken, te weten Biologie, Frans en Geschiedenis.

³ Bij VWO'ers oude stijl wordt het gemiddelde van de maximaal zeven hoogste cijfers genomen.

3.2 Het cijfermateriaal

De Universiteit Leiden beschikt over de eindresultaten voor de eindexamenvakken, en het profiel dat de leerling heeft gekozen. De eindcijfers van het diploma worden door de studenten bij eerste inschrijving op het inschrijfformulier genoteerd en ingevoerd. Ook geven zij hun vooropleiding aan; voor de leerlingen nieuwe stijl wil dat zeggen het profiel of de combinatie van profielen. De gegevens worden vervolgens van het formulier overgenomen, waarbij steeds de vakcodes van de vakken moeten worden ingetypt; deze codes staan naast de namen van de desbetreffende vakken op het formulier.

De geregistreerde cijfers zijn niet foutloos. Er zijn drie mogelijke foutenbronnen:

- De gegevens staan onduidelijk of niet goed vermeld op het diploma⁴
- De student neemt de gegevens verkeerd over op het inschrijfformulier⁵,
- De administratie maakt een fout bij het invoeren van de gegevens⁶.

Voor een deelgroep van onze onderzoeksgroep bleek het mogelijk de dossiers te lichten en opnieuw VWO-cijfers in te voeren. Het ging om de deelgroep van 811 studenten met een VWO-diploma nieuwe stijl die zich in 2003 voor het eerst aan de Universiteit Leiden inschreven, voor de studie Geschiedenis, Rechten of Psychologie. De invoer vond plaats via een formulier in Access; in plaats van het intypen van een vakcode, kon nu door de invoerder worden volstaan met het kiezen van het desbetreffende vak uit een menu. Het idee was dat deze wijze van invoer nauwkeuriger was⁷. Tegelijk werd van de gelegenheid gebruik gemaakt om de cijfers van het centraal examen vast te leggen en om vast te leggen of een vak in de keuzeruimte of het profiel thuis hoorde. Uit een vergelijking van de gegevens uit de

⁴ De opgave en codering voor maatschappijleer is lastig. Maatschappijleer in het algemeen gedeelte wordt dikwijls als *maatschappijleer* opgegeven. Ter onderscheiding van het vak maatschappijleer in het keuzedeel, een vak met een centraal examenonderdeel, is het beter om dit vak met *maatschappijleer 1* aan te geven. Op het diploma staan soms ook echte fouten. Zo wordt een vak als Frans 1,2 op het diploma een enkele keer ten onrechte onder het algemeen deel vermeld.

⁵ Er zijn veel (59) vakken, met labels als “Mijleer”, “Gesch 1”, “Geschiedenis”, “Gesch/Mijleer”, “Mijleer 1” en deze labels staan niet op alfabet, maar op volgorde van de code waaronder de gegevens opgeslagen dienen te worden.

⁶ Wellicht kan in de toekomst na toestemming gebruik gemaakt worden van gegevens die centraal opgeslagen worden. Zolang dat niet het geval is, blijft invoering van de gegevens handwerk.

⁷ Inconsistenties kunnen hierbij voor een gedeelte softwarematig worden voorkomen (door inperking van de keuzemenu's in bijvoorbeeld een Access-formulier) of achteraf worden opgespoord en gecorrigeerd. Er heeft in dit onderzoek een beperkte consistentiecheck achteraf plaatsgevonden.

Access-database en de cijfers uit het centrale gegevensbestand achteraf, waarbij ook enkele diploma's opnieuw werden bekeken bleek dat alle genoemde soorten fouten voorkwamen⁸.

3.3 Het centraal schriftelijk versus het eindcijfer

Aangezien wij voor een deelgroep van studenten beschikten over het eindcijfer én het cijfer voor het centraal schriftelijk, hebben wij de relatie tussen deze cijfers onderzocht. In Tabel 1 staat het gemiddelde eindcijfer en het gemiddelde voor het centraal schriftelijk voor de 23 vakken die door meer dan 25 studenten waren gedaan⁹.

Tabel 1. Gemiddelde cijfers en aantallen voor 23 eindexamenonderdelen

	eindcijfer		centraal schrift.
	gemiddelde	gemiddelde	aantal
Nederlands	6.9	6.7	810
Latijn	6.6	6.2	176
Grieks	6.7	6.3	70
Frans 1,2	6.6	6.3	293
Duits 1,2	6.7	6.5	277
Engels	6.9	6.7	810
Geschiedenis	6.9	6.7	656
Aardrijkskunde	6.6	6.4	468
Wiskunde A1	6.5	6.5	241
Wiskunde A1,2	6.4	6.1	367
Wiskunde B1	6.2	6.1	151
Wiskunde B1,2	6.4	6.5	55
Natuurkunde 1	6.4	6.3	138
Natuurkunde 1,2	6.7	6.8	43
Scheikunde 1	6.4	6.3	123
Scheikunde 1,2	6.8	6.8	52
Biologie 1,2	6.4	6.2	188
Economie 1	6.2	5.8	147
Economie 1,2	6.4	6.1	375
Management en organisatie	6.4	6.2	194
Tekenen	7.2	6.9	75
Filosofie	7.1	6.8	89
Maatschappijleer	7.0	6.7	42
Totaal	6.7	6.4	5840

⁸ In het Access-bestand is maar 44 keer *maatschappijleer* gecodeerd als keuzevak met een centraal schriftelijk examendeel; een groot aantal keren is *maatschappijleer* op het diploma en/of in het universitaire gegevensbestand in het accesbestand gecodeerd als maatschappijleer 1.

⁹ In de tabel zijn alleen gegevens opgenomen van studenten voor wie zowel het cijfer voor het centrale deel als het eindcijfer in de database waren opgenomen. Bij één student ontbrak het cijfer voor het centraal schriftelijk voor de voor iedereen verplichte vakken Nederlands en Engels. Daardoor komt het grootste aantal op 810 uit i.p.v. 811.

De gemiddeldes voor de eindcijfers lopen uiteen van 7.2 voor *Tekenen* tot 6.2 voor *Economie 1*; de gemiddeldes voor het centraal schriftelijk variëren van een relatief hoog gemiddelde van 7 bij *Tekenen* tot een lage 5.8 bij *Economie 1*. Een deel van de verschillen in geobserveerde moeilijkheid is te wijten aan verschillen tussen de groepen studenten die de desbetreffende vakken in hun pakket hebben genomen, een ander deel van de verschillen is te wijten aan het verschillen tussen de vakken.

Het centraal schriftelijk ligt gemiddeld ongeveer tertiële punt onder het eindcijfer. Dit betekent dat het verschil in gemiddeldes tussen het centraal schriftelijk en het schoolexamen nog groter is. De omvang van de discrepantie verschilt van vak tot vak. Bij Natuurkunde1,2 is het gemiddelde van het centraal schriftelijk zelfs iets hoger dan het gemiddelde eindexamencijfer, maar dit kan op toeval berusten. De verwachte grootte van het verschil tussen het centraal schriftelijk en het schoolonderzoek hangt af van het niveau van de leerling. Bij leerlingen die heel hoog scoren bij het centraal schriftelijk mag men geen hoger cijfer voor het schoolonderzoek verwachten (tegenover een 10 op het centraal schriftelijk kan niet eens een hoger cijfer voor het schoolonderzoek voorkomen).

De samenhang tussen het centraal schriftelijk en het eindexamencijfer is in het algemeen redelijk. De hoogste correlatie wordt gevonden voor Natuurkunde1,2, namelijk 0.92. De laagste correlatie is die bij Nederlands: 0.75. Deze correlatie is beslist laag te noemen, als wij bedenken dat het centraal schriftelijk voor de helft in het eindcijfer meetelt¹⁰.

Verschillende factoren spelen een rol bij het ontstaan van verschillen tussen schoolonderzoek en centraal examen. Toeval speelt zeker een rol, maar ook het feit dat de beoordeling bij het schoolonderzoek per definitie niet gestandaardiseerd is. Ten slotte, bij het schoolonderzoek worden deels andere kennis en vaardigheden getoetst dan bij het centraal schriftelijk. Sommige vaardigheden zijn moeilijk centraal te toetsen.

¹⁰ Bij de interpretatie van de hoogte van de correlatie moet ook rekening worden gehouden met het feit dat het eindexamencijfer ontstaat na afronding van het resultaat dat verkregen wordt door het centraal schriftelijk en het schoolexamencijfer te middelen. De samenhang tussen schriftelijk en schoolexamen is geringer. Deze samenhang zou nauwkeuriger moeten worden onderzocht in een onderzoek waarbij ook het cijfer van het schoolexamen wordt vastgelegd.

Er is uitgebreider onderzoek verricht naar schoolonderzoeken en centrale examens. Een goed overzicht biedt het werk van Van den Bergh, Rohde en Zwarts (2003). Zij verwijzen onder meer naar een onderzoek van Pijl (1991), waaruit blijkt dat scholen m.b.t. de cijfergeving bij het schoolonderzoek systematisch verschillen en dat deze verschillen niet verklaard kunnen worden door verschillen tussen leerlingen. Het resultaat is eigenlijk niet zo verbazingwekkend: verschillende scholen hebben verschillende manieren om het schoolonderzoek te verrichten, en bij de beoordeling van het werk van de leerlingen biedt de prestatie van de 'gemiddelde leerling' bij gebrek aan een objectief ijkpunt houvast. Dat maakt het schoolonderzoek een zwakke schakel in een selectieprocedure. Het centrale examen aan het eind van het Nederlandse voortgezet onderwijs waarborgt het niveau van de eisen die aan de leerlingen worden gesteld, zo heeft ook de onderwijsraad (Onderwijsraad, 2005) in een recent rapport opgemerkt.

Anderzijds blijkt uit onderzoek naar onderzoek van Schooten en De Glopper (2002) dat de normering van de centrale examens evenmin zonder problemen is. Van den Bergh *et al.* laten zien dat de cijfers van het centraal examen over de jaren heen niet strikt vergelijkbaar zijn, ondanks de toepassing van normhandhavingsonderzoek door het CITO. Eén van de redenen dat de cijfers voor het centraal schriftelijk in de periode waarop het onderzoek betrekking had niet 'stabiel' waren, is, zo melden Van den Bergh *et al.* het feit dat de definitieve cijfers niet alleen gebaseerd waren op de resultaten van het normhandhavingsonderzoek.

3.4 De verplichte onderdelen en de keuzeruimte bij een deelgroep

De gegevens die in Access zijn ingevoerd, zijn verder geanalyseerd. Enkele beschrijvende gegevens m.b.t. de gegevens voor de deelgroep van 811 studenten worden in deze paragraaf gepresenteerd. Het aantal onderdelen uit het algemene gedeelte en het profieldeel dat met een cijfer (i.p.v. een beoordeling) is afgesloten varieert van 7 tot 14. De modus ligt bij 12 onderdelen. Het aantal keuzeonderdelen met een cijfer varieert van 0 tot 5, met de modus bij 1. Het aantal keuzevakken (afgesloten met een cijfer) correleert niet met het aantal verplichte vakken met een cijfer. Het aantal keuzevakken correleert licht positief ($r = 0.31$) met het gemiddelde cijfer voor de verplichte onderdelen, en nauwelijks met het gemiddelde cijfer voor de keuzeonderdelen ($r = 0.20$). Het gemiddelde cijfer waarbij van de keuzevakken alleen het hoogste cijfer wordt genomen, de basis voor de indeling in

lotingsklassen bij numerus fixus opleidingen, verschilt pas in de tweede decimaal van het gemiddelde, berekend over alle eindexamenvakken met een cijfer. De correlatie tussen de twee berekeningen van het gemiddelde cijfer is heel hoog ($r = 0.996$). Dat betekent dat bij het gemiddelde eindexamencijfer in berekeningen een goede benadering zou zijn van de berekening van het gemiddelde zoals de IB-groep dat doet. Als we ons beperken tot het gemiddelde eindcijfer van de in Tabel 1 genoemde vakken en enkele andere vakken met een centraal schriftelijk eindexamen (Spaans, Russisch, Turks, Muziek, Handvaardigheid), dan krijgen wij een andere indicator, het gemiddelde eindcijfer voor alle vakken die ook een centraal schriftelijk examengedeelte hebben, die heel hoog met het gemiddelde cijfer volgens de IB-groep correleert ($r = 0.93$). Het gemiddelde van de cijfers op de centrale toetsing correleert 0.83 met het gemiddelde volgens de IB-groep.

4. De cohorten

Van alle in het onderzoek betrokken opleidingen werd voor de jaren 2001/2002, 2002/2003 en 2003/2004 nagegaan hoeveel voltijdstudenten voor het eerst voor de propedeusefase van de studie waren ingeschreven. In Tabel 2 staan deze aantallen vermeld. De complexiteit van de selecties die gepleegd moeten worden houdt in dat de aantallen uit Tabel 2 af kunnen wijken van de aantallen studenten waar de opleidingen zelf van uit gaan¹¹. Voor het huidige onderzoek is dat geen probleem.

Tabel 2. *De omvang van de cohorten voltijd studenten*

	2001	2002	2003
Rechten	728	673	746
Bestuurskunde	167	101	115
Psychologie	453	455	558
Frans	26	29	24
Italiaans	27	13	23
Duits	10	10	15
Geschiedenis	137	193	187

In de cohorten zitten studenten die niet op basis van een VWO-diploma aan de studie begonnen zijn. Voor het huidige onderzoek moeten wij alleen kijken naar studenten van wie eindexamencijfers bekend zijn. De omvang van de uit deze selectie resulterende groepen staat in Tabel 3.

Tabel 3. *De omvang van de deelgroepen met VWO-cijfers*

	2001	2002	2003
Rechten	526	477	558
Bestuurskunde	69	89	86
Psychologie	343	343	419
Frans	23	22	13
Italiaans	21	7	17
Duits	5	6	11
Geschiedenis	114	153	141

Voor het onderzoek zijn voor alle opleidingen, uitgezonderd Psychologie, de groepen uit Tabel 3 het uitgangspunt voor het onderzoek. Bij Psychologie wordt

¹¹ In de centrale selectie ontbreken de zogenaamde “bijvakkers”.

uitgegaan van de eigen cohortdefinitie, die betrekking heeft op de studie jaren 2002/2003 en 2003/2004.

Van de studenten uit Tabel 3 is nog een deelgroep van belang: de studenten die het VWO nieuwe stijl hebben gevolgd. Van deze deelgroep staan de aantallen studenten in Tabel 4. In de periode waarover dit onderzoek betrekking heeft, is de relatieve instroom van VWO'ers nieuwe stijl van praktisch 0 (6 procent) naar 80 procent gestegen.

Tabel 4. *De percentages VWO'ers nieuwe stijl*

	2001	2002	2003
Rechten	6.1	61.2	84.9
Bestuurskunde	5.8	58.4	75.6
Psychologie	5.2	55.1	76.6
Frans	8.7	22.7	84.6
Italiaans	.0	71.4	58.8
Duits	.0	66.7	63.6
Geschiedenis	5.3	38.6	73.0

De verdeling over de profielen verschilt van opleiding tot opleiding. De verdeling van de studenten over de profielen staat in Tabel 5. Het percentage studenten met het profiel *Natuur & Techniek* is uiterst gering, evenals het percentage studenten met meer dan één profiel. Bij Rechten en Bestuurskunde zijn er relatief veel studenten met een *Economie & Maatschappij* profiel. Bij Psychologie heeft een groot percentage studenten een *Cultuur & Maatschappij* achtergrond, maar er zijn ook relatief veel studenten met het profiel *Natuur en Gezondheid* of het profiel *Economie en Maatschappij*. Bij de drie talenstudies komt, zoals verwacht, vooral het profiel *Cultuur & Maatschappij* voor. Bij Geschiedenis komt naast *Cultuur & Maatschappij* het profiel *Economie & Maatschappij* relatief vaak voor.

Tabel 5. *Verdeling over de profielen*

	N&T	N&G	E&M	C&M	N&T- N&G	E&M- C&M
Rechten	3.0	11.2	56.6	27.9	.6	.6
Bestuurskunde	1.7	10.7	66.9	18.2	1.7	.8
Psychologie	4.5	26.9	27.5	40.0	.8	.4
Frans		5.6	11.1	77.8		5.6
Italiaans		20.0	6.7	73.3		
Duits			27.3	72.7		
Geschiedenis	6.0	7.7	38.1	48.2		
Totaal	3.6	15.7	45.1	34.4	.7	.5

In hoofdstuk 3 werd er melding van gemaakt dat van een deel van de VWO'ers nieuwe stijl uit 2003/2004 de cijfers van het eindexamen opnieuw waren ingevoerd. Er bleken 427 Rechten studenten voor te komen, 299 Psychologie studenten en 89 Geschiedenis studenten.¹² De gemiddelde cijfers voor deze studenten voor de meest voorkomende VWO-vakken worden gegeven in Tabel 6. Geschiedenis lijkt studenten aan te trekken die gemiddeld beter in hun talen en Geschiedenis zijn. De verschillen tussen de opleidingen voor VWO-Geschiedenis en Frans^{1,2} zijn significant.

Tabel 6. *Gemiddelde eindexamencijfers voor 815 studenten. Bron: Accessbestand*

	Rechten	Psychologie	Geschiedenis
Nederlands	6.9	6.9	7.0
Latijn	6.5	6.5	7.0
Frans 1,2	6.6	6.4	6.9
Duits 1,2	6.6	6.7	6.8
Engels	6.9	6.8	7.2
Geschiedenis	6.9	6.7	7.3
Aardrijkskunde	6.6	6.5	6.9
Wiskunde A1	6.5	6.6	6.4
Wiskunde A1,2	6.4	6.5	6.0
Biologie 1,2	6.4	6.5	6.3
Economie 1,2	6.4	6.3	6.6
Management en organisatie	6.5	6.2	6.4

¹² Dat is totaal 815. Er waren gegevens van 811 studenten ingevoerd; enkele studenten komen bij twee studies voor. Zij doen twee studies of zijn al in een vroeg stadium omgezwaid.

5. **De relatie tussen gemiddelde eindexamencijfers en het al dan niet behalen van de propedeuse na 1, resp. 2 jaar**

In dit hoofdstuk gaan wij voor VWO'ers de voorspelling van het studiesucces in termen van het al dan niet behalen van de propedeuse binnen 1, resp. 2 jaar na. De voorspelling is gebaseerd op gemiddelde eindexamencijfers. Er wordt gewerkt met verschillende voorspellers op basis van eindexamencijfers. Met name wordt gewerkt met de definitie die bij gewogen loting wordt gebruikt¹³, en waarbij de cijfers worden getransformeerd naar een aantal lotingsklassen. Ondanks het feit dat wij ons in dit onderzoek niet met een toepassing van gewogen loting voor een opleiding bezig houden, zullen wij bij de berekeningen waar berekening volgens het systeem van gewogen loting wordt gebruikt, over *lotingsklassen* schrijven. De cohorten zijn op centraal niveau gedefinieerd met uitzondering van de cohorten Psychologie. Voor de cohorten Psychologie en de resultaten bij Psychologie gebruiken wij in dit hoofdstuk de gegevens van een eerder onderzoek dat reeds werd afgesloten (De Gruijter, 2004). Wij beginnen met een verslag van de uitkomsten bij Psychologie.

5.1. **Psychologie; cohorten 2002 en 2003**

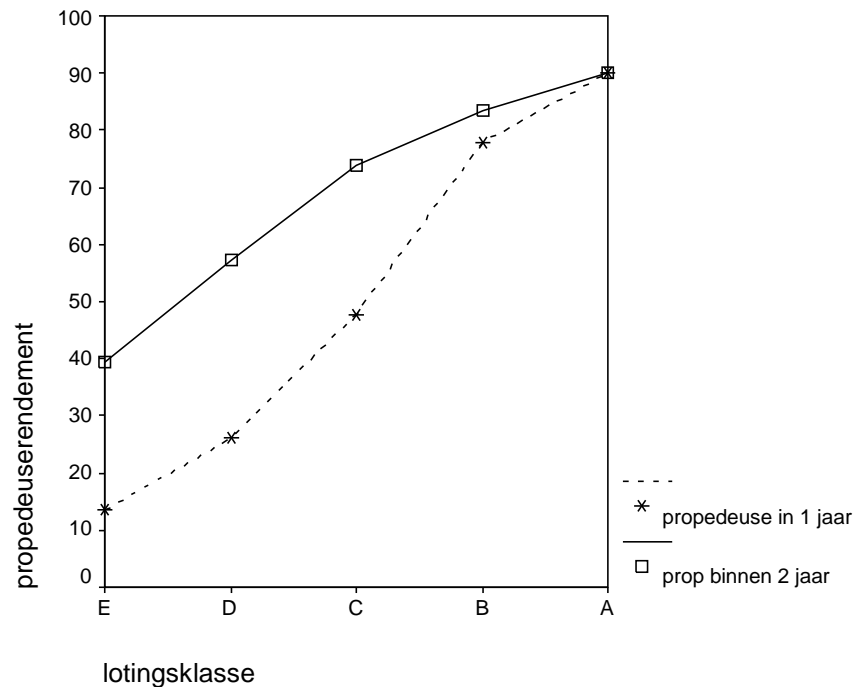
Bij Psychologie werd in het kader van de invoering van de numerus fixus vanaf het studiejaar 2004-2005 al een onderzoek gedaan naar de samenstelling van de voltijd cohorten 2002, 2003 en 2004, en de samenhang tussen vooropleidingresultaten en studiesucces, gedefinieerd als geslaagd voor de propedeuse binnen één, resp. twee jaar. Op basis van de in het universitaire registratiesysteem vastgelegde vooropleidinggegevens werd van elke student een schatting van de lotingsklasse gemaakt. De resultaten van het onderzoek naar de samenhang tussen lotingsklasse en studiesucces vatten wij samen in Tabel 7.

¹³ Voor het profiel *Cultuur en Maatschappij*, gekenmerkt door vrijheid in de keuze van vakken, is het gemiddelde volgens de IB-groep niet exact te repliceren vanuit de in het studentregistratiesysteem beschikbare gegevens, aangezien in dat systeem niet is vastgelegd welke vakken keuzevakken zijn en welke vakken bij het profieldeel horen.

Tabel 7. De relatie tussen lotingsklasse en het al dan niet behalen van de propedeuse in één of twee jaar

		2002		2003		2002	
		propedeuse in 1 jaar		propedeuse in 1 jaar		propedeuse in 2 jaar	
		%	Aantal in klasse	%	Aantal in klasse	%	Aantal in klasse
lotingsklasse	C*	17.7	113	17.8	146	43.4	113
	E	13.4	119	11.4	123	39.5	119
	D	26.1	119	25.6	156	57.1	119
	C	47.8	69	37.1	97	73.9	69
	B	77.8	18	75.0	28	83.3	18
	A	90.0	10	64.3	14	90.0	10
Totaal		27.5	448	25.9	564	53.3	448

In de tabel maken wij onderscheid tussen 6 prestatieniveau's. De niveau's A t/m E volgen de indeling van de IB-groep naar lotingsklasse op basis van gemiddeld eindexamencijfer. Groep C* bevat de studenten die in lotingsklasse C vallen op basis van een ander diploma dan dat van het VWO; deze groep hebben wij voor het onderzoek dus apart gehouden. Het percentage studenten dat na 1, respectievelijk 2 jaar de propedeuse heeft behaald, neemt duidelijk toe met de lotingsklasse, als wij van lotingsklasse E naar lotingsklasse A gaan (als wij de lotingsklassen de cijfers 1 t/m 5 toekennen, bedraagt de samenhang met het dichotome criterium *na twee jaar al dan niet geslaagd voor de propedeuse* in termen van de product-moment correlatie 0.31). De samenhang tussen propedeuserendement en lotingsklasse is grafisch weergegeven in Figuur 1. De groep studenten in lotingsklasse C* blijkt lager te presteren dan de VWO'ers uit lotingsklasse C. Op dit verschil gaan wij hier niet in: onze belangstelling gaat in dit onderzoek uit naar de groep studenten van wie wij VWO-cijfers hebben.



Figuur 1. *Het propedeuserendement voor de cohort 2002.*

Het rendement na 2 jaar bij Psychologie is 53 procent (op basis van de gegevens van de cohort 2002). Laten wij alle studenten toe uit de categorieën waarbinnen het studiesucces na twee jaar groter is dan 50 procent, dan stijgt het rendement, ceteris paribus, tot boven de 65 procent. Als wij alleen studenten uit de hogere lotingsklassen A, B en C tot de studie toelaten, stijgt het rendement zelfs tot boven de 75 procent. Uit Tabel 2 blijkt nu dat de hoogste lotingsklassen klein zijn. Als wij alleen studenten uit de lotingsklassen A, B en C zouden toelaten, dan zouden wij een selectieratio van 22 procent moeten hanteren. M.a.w., wij zouden slechts 22 procent van de studenten die zich aanmelden, moeten accepteren.

Bij Psychologie werd het onderzoek verricht in het kader van de invoering van een numerus fixus die nodig was omdat de opleiding de wassende stroom van studenten niet meer kon verwerken. De opleiding maakt vanaf het studiejaar 2004 gebruik van het systeem van gewogen loting voor de instroomselectie. De verwachte selectieratio bij psychologie ligt eerder in de buurt van de 80 procent dan in de buurt van de 20 procent.¹⁴ De rendementswinst bij een dergelijke selectieratio slechts enkele percentagepunten.

¹⁴ Invoering van het systeem resulteerde in 2004 zelfs in een veel lagere instroom dan de numerus fixus.

Bij Psychologie werd ook naar de relatie tussen het cijfer voor Wiskunde en studiesucces bekeken. Uit dit onderzoek bleek dat de voorspelling van studiesucces iets beter werd indien met het cijfer voor wiskunde rekening zou worden gehouden. Het belang van enig wiskundig begrip wordt door de opleiding erkend en heeft al geresulteerd in een wiskunde-eis aan aankomende studenten met een wiskundedeficiëntie.

Ten behoeve van het huidige onderzoek zijn enkele extra berekeningen uitgevoerd, waarbij gebruik is gemaakt van de cijfers die in de Accessdatabase waren opgeslagen. Van 299 studenten uit de cohort 2003 waren cijfers uit deze database beschikbaar. Voor deze deelgroep konden enkele verschillende voorspellers van studiesucces worden berekend:

- het gemiddelde van de verplichte cijfers samen met het hoogste cijfer voor een keuzevak; dit cijfer geeft na indeling in de categorieën A t/m E de classificatie volgens de IB-groep
- het gemiddeld over alle cijfers
- het gemiddelde over de cijfers van de centrale toetsen
- het gemiddelde van de eindcijfers voor die vakken waarvoor ook een cijfer voor de centrale toets beschikbaar was.

De samenhang tussen deze voorspellers en het voorlopige studiesucces (het al dan niet behalen van de propedeuse na één jaar, wordt gegeven in Tabel 8. De correlaties in de tabel ontlopen elkaar uiteraard niet veel. Op de waarde van de schoolexamens is kritiek mogelijk: de vergelijking tussen leerlingen van verschillende scholen kan problematisch zijn vanwege verschillen in de accenten die de verschillende scholen leggen en de verschillende normen. Een voorspelling gebaseerd op alleen de cijfers voor de centrale toetsen is daarom een a-priori aantrekkelijke mogelijkheid. In dit onderzoek blijkt deze voorspeller de geringste samenhang met het criterium al dan niet halen van de propedeuse te vertonen ($r = 0.36$). De hoogste correlatie ($r = 0.40$) vindt men bij een voorspelling met gemiddelde eindcijfers gebaseerd op examenonderdelen met een centrale toetsing en een schoolexamen. Het is mogelijk dat de schoolexamens die deels andere vaardigheden meten dan de centrale toetsen, en die een langere periode van toetsing betreffen, daardoor extra valide informatie bevatten die compenseren voor de extra ruis die inherent is aan een schoolonderzoek. Ook kan een rol spelen dat het gemiddelde op basis van alle vakken met een centraal cijfer gebaseerd is op

tweemaal zoveel cijfers als het gemiddelde gebaseerd op de centrale cijfers alleen en dat geeft, ondanks een grotere invloed van foutenbronnen bij het schoolexamen, de doorslag. De verschillen tussen de correlaties zijn overigens klein. Men mag in ieder geval concluderen dat in dit onderzoek de voorspelling op basis van uitslagen van centrale toetsen niet superieur is aan de andere voorspellingen.

Tabel 8. *De correlatie tussen propedeuse succes en verschillende voorspellers voor een deelgroep van 299 studenten uit de cohort 2003*

	propedeuse in 1 jaar
gem. cijfer met 1 keuzevak	.390
gemiddelde alle cijfers	.405
gemiddelde centrale toetsen	.357
gemiddelde eindcijfer van alle vakken met een centraal cijfer	.410

De vraag is al gerezen of de voorspelling op basis van eindexamencijfers niet lastig is vanwege het feit dat er zo weinig vakken zijn die door alle leerlingen ook met een centraal examengedeelte worden afgesloten. Met name is er een verschil tussen *N&T* en *N&G* enerzijds en *C&M* en *E&M* anderzijds. Om dat na te gaan is onderzocht of de voorspelling op basis van het gemiddelde eindexamencijfer van alle vakken met een centraal cijfer kan worden verbeterd door rekening te houden met het profiel van de leerling. De groep studenten werd daartoe in twee deelgroepen gesplitst: studenten met een profiel *N&T* en/of *N&G* (102 studenten) versus de overige studenten (197 studenten). Er werd een multiple regressie analyse gedaan met gemiddeld cijfer en deelgroep als voorspellers. De informatie over het profiel dat de student had gevolgd, bleek geen significante bijdrage aan de voorspelling van studiesucces in termen van het al dan niet halen van de propedeuse binnen één jaar te leveren. Ook kon met de kleinere aantallen gegevens uit de Accessdatabase de significante verbetering van de voorspelbaarheid van studiesucces (nu in termen van het behalen van de propedeuse na één jaar) niet gerepliceerd worden.

5.2. Bestuurskunde; cohorten 2001, 2002 en 2003

Bij Bestuurskunde zijn de cohorten die onderzocht worden, veel kleiner dan de cohorten bij Psychologie: 69 studenten in de cohort 2001, 89 in de cohort 2002 en 86 in de cohort 2003 (zie Tabel 3). De geringere omvang van de cohorten maakt

het moeilijker om effecten te vinden, zoals nog zal blijken. Wij hebben voor alle studenten het gemiddeld cijfer van de VWO-vakken berekend volgens de richtlijnen bij de procedure voor gewogen loting¹⁵ en afgerond tot de lotingscategorieën A t/m E. In Tabel 9 staat de verdeling per cohort over de verschillende categorieën. De hoogste categorie, categorie A, komt heel weinig voor. De grootste groep studenten is groep D, de groep met gemiddelde cijfers tussen de 7 en de 6.5.

Tabel 9. *De verdeling van de studenten over lotingsklassen bij Bestuurskunde*

		2001	2002	2003	Totaal
lotingsklasse	E	26	20	19	65
	D	23	29	36	88
	C	13	21	15	49
	B	3	17	13	33
	A	4	2	3	9

De percentages geslaagden na twee jaar voor elk van de categorieën staan, uitgesplitst naar cohort, in Tabel 10. Het percentage geslaagden voor de hoogste categorie valt tegen; daarbij moet men wel bedenken dat dit percentage op slechts 6 studenten is gebaseerd. Als wij categorie A buiten beschouwing laten, valt op dat het percentage na twee jaar geslaagde studenten alleen bij de laagste categorie onder de 50 procent valt. Voor alle drie de onderzochte cohorten beschikken wij over het percentage studenten dat binnen een jaar voor de propedeuse slaagt. Deze percentages staan, uitgesplitst naar lotingscategorie in Tabel 11. In de cohort 2003-2004 zien wij bij dit criterium een sterk verschil tussen de twee hoogste en de twee laagste categorieën.

Tabel 10. *Het percentage binnen twee jaar geslaagde studenten Bestuurskunde*

		2001	2002	Totaal
lotingsklasse	E	34.6	30.0	32.6
	D	65.2	51.7	57.7
	C	61.5	61.9	61.8
	B	100.0	58.8	65.0
	A	25.0	50.0	33.3

¹⁵ Zie voetnoot 13.

Tabel 11. *Het percentage binnen één jaar geslaagde studenten Bestuurskunde*

		2001	2002	2003	Totaal
lotingsklasse	E	23.1	5.0	10.5	13.8
	D	39.1	27.6	8.3	22.7
	C	53.8	47.6	33.3	44.9
	B	33.3	52.9	53.8	51.5
	A	25.0	50.0	66.7	44.4

Zoals reeds vermeld, doen relatief veel Bestuurskunde studenten het profiel *Economie en Maatschappij*. Relatief veel studenten met VWO nieuwe stijl hebben dan ook een cijfer voor Economie 1,2. Ook een vak als Management en Organisatie komt voor. Aangezien deze twee vakken als nuttig voor het volgen van Bestuurskunde werden beschouwd, is ook gekeken naar de samenhang tussen de cijfers voor deze vakken en het al dan niet na één, respectievelijk twee jaar behaald hebben van de propedeuse in de cohorten 2002 en 2003. Bij de cohort 2002 werden geen significante relaties tussen de cijfers voor deze vakken en studiesucces gevonden; het aantal studenten dat Management en Organisatie had gedaan, was ook wel erg klein (N = 17). In de cohort 2003 bleek Management en Organisatie een vergelijkbare samenhang met studiesucces te vertonen als de lotingsklasse, maar de correlatie is slechts gebaseerd op de gegevens van 22 studenten.

5.3. Geschiedenis; cohorten 2001, 2002 en 2003

Bij Geschiedenis was de omvang van de cohorten 114 voor de cohort 2001, 153 voor de cohort 2002 en 141 voor de cohort 2003. De verdeling over de lotingsklassen wordt gegeven in Tabel 12. De meeste studenten vallen in de categorieën C en D.

Tabel 12. *De verdeling van de studenten over lotingsklassen bij Geschiedenis*

		2001	2002	2003	Totaal
lotingsklasse	E	29	41	28	98
	D	36	47	48	131
	C	23	43	37	103
	B	13	12	20	45
	A	13	10	8	31

De percentages studenten die na twee jaar geslaagd zijn voor de propedeuse, staan uitgesplitst naar lotingsklasse in Tabel 13, de percentages na één jaar geslaagden staan in Tabel 14. De percentages lopen duidelijk met de lotingsklasse op.

Tabel 13. *Het percentage binnen twee jaar geslaagde studenten Geschiedenis*

		2001	2002	Totaal
lotingsklasse	E	24.1	34.1	30.0
	D	41.7	38.3	39.8
	C	65.2	58.1	60.6
	B	92.3	83.3	88.0
	A	92.3	90.0	91.3

Tabel 14. *Het percentage binnen één jaar geslaagde studenten Geschiedenis*

		2001	2002	2003	Totaal
lotingsklasse	E	3.4	7.3	10.7	7.1
	D	16.7	21.3	12.5	16.8
	C	30.4	30.2	27.0	29.1
	B	69.2	66.7	40.0	55.6
	A	92.3	90.0	75.0	87.1

Voor de deelgroep van studenten van wie de gegevens in de Access database waren ingevoerd, zijn extra analyses gedaan. Het gaat hierbij om 89 studenten uit de cohort 2003. Op één student na hadden alle studenten Geschiedenis in hun eindexamenpakket, de meesten binnen hun profiel (C&M of E&M). Bij de opleiding Geschiedenis is het dus heel nuttig te kijken naar de samenhang met het verwante vak in het VWO. Eerst kijken wij naar de relatie tussen het al dan niet in één jaar behalen van de propedeuse en gemiddelde cijfers. Net zoals bij Psychologie werden verschillende gemiddelden berekend. De hoogste correlaties met het criterium werden gevonden voor de twee gemiddelden berekend over de vakken die een centraal onderdeel kennen. Van deze twee gemiddelden gaf, zoals bij Psychologie het geval was, het gemiddelde van de eindcijfers het beste resultaat: de correlatie van dit gemiddelde met het criterium was 0.52. Het vak Geschiedenis correleert 0.41 met het criterium, 0.67 met het gemiddelde eindcijfer. De correlatie van Geschiedenis met het criterium is vrij hoog. Enerzijds mag men verwachten dat gemeenschappelijke aspecten, die o.a. eenzelfde beroep op de motivatie van studenten, respectievelijk leerlingen doen, tot een relatief grote samenhang kunnen leiden. Anderzijds moet men in dit soort situaties beducht zijn

voor zelfselectie-effecten. Dat wil zeggen, studenten die niet gemotiveerd zijn voor Geschiedenis, en daardoor lagere cijfers voor het VWO-vak halen, kiezen geschiedenis niet als vervolgopleiding. Het gevolg hiervan is dat de range van cijfers voor het VWO-vak beperkt wordt en de samenhang tussen de VWO-cijfers en resultaten in de vervolgopleiding in termen van de correlatiecoëfficiënt lager uitvalt dan men in eerste instantie zou verwachten. Het is dan ook niet verwonderlijk dat toevoeging van het VWO-cijfer Geschiedenis als voorspeller geen verbeterde voorspelling van studiesucces oplevert.

5.4. Rechten; cohorten 2001, 2002 en 2003

Rechten heeft van alle onderzochte opleidingen de grootste aantallen studenten. De cohort 2001 heeft 526 studenten, de cohort 2002 477, en de cohort 2003 558 (zie tabel 3). De verdeling over de lotingsklassen wordt gegeven in Tabel 15. De meeste studenten vallen in de lagere categorieën E en D.

Tabel 15. De verdeling van de studenten over lotingsklassen bij Rechten

		2001	2002	2003	Totaal
lotingsklasse	E	211	170	162	543
	D	153	155	182	490
	C	102	97	142	341
	B	37	39	49	125
	A	23	16	23	62

In Tabel 16 staan de gegevens m.b.t. het criterium *propedeuse gehaald binnen twee jaar*. De trend in de percentages is duidelijk. Er is één uitzondering: het percentage geslaagden in klasse A is relatief laag in 2002. Dit percentage is gebaseerd op de prestaties van slechts 16 studenten.

Tabel 16. Het percentage binnen twee jaar geslaagde studenten Rechten

		2001	2002	Totaal
lotingsklasse	E	47.4	34.1	41.5
	D	66.7	43.2	54.9
	C	77.5	72.2	74.9
	B	83.8	87.2	85.5
	A	91.3	68.8	82.1

Tabel 17 geeft voor alledrie de cohorten de percentages geslaagden na één jaar, uitgesplitst naar lotingsklasse.

Tabel 17. *Het percentage binnen één jaar geslaagde studenten Rechten*

		2001	2002	2003	Totaal
lotingsklasse	E	18.0	12.9	13.6	15.1
	D	32.0	21.3	26.9	26.7
	C	56.9	54.6	54.2	55.1
	B	54.1	64.1	67.3	62.4
	A	73.9	62.5	69.6	69.4

Voor de deelgroep van studenten van wie de gegevens in de Access database waren ingevoerd, zijn extra analyses gedaan. Van 427 studenten uit de cohort 2003 waren extra gegevens bekend. Bij deze groep werden verschillende gemiddelden berekend. De correlatie van het gemiddelde gebaseerd op de cijfers van de centrale toetsen, met het criterium *al dan niet voor de propedeuse geslaagd na één jaar* was het laagst: 0.41. De andere correlaties ontliepen elkaar nauwelijks. De hoogste correlatie werd gevonden bij het gemiddelde gebaseerd op de eindcijfers van de vakken die zowel een centraal als een schoolexamen kennen: 0.46.

5.5. Drie talenstudies

Voor de jaren waar dit onderzoek betrekking op heeft, hebben wij VWO-cijfers van 58 eerstejaars Frans, 45 eerstejaars Italiaans en 22 eerstejaars Duits. De verdeling van de studenten over de verschillende lotingsklassen wordt gegeven in Tabel 18.

Tabel 18. *De verdeling van de studenten over lotingsklassen bij drie talenstudies*

		Frans	Italiaans	Duits
lotingsklasse	E	15	8	4
	D	9	15	5
	C	24	13	5
	B	6	4	5
	A	4	5	3
Totaal		58	45	22

Voor de studenten uit de cohorten 2001 en 2002 werd het propedeuserendement na twee jaar per lotingsklasse berekend. De desbetreffende gegevens staan in Tabel 19. Met name de gegevens van Duits, gebaseerd op slechts 11 studenten (zie Tabel 3) zijn erg gevoelig voor toevalseffecten. De gegevens in Tabel 20 m.b.t. tot het al dan niet halen van de propedeuse in één jaar zijn gebaseerd op de gegevens van alle cohorten; ook hier blijft het aantal studenten Duits gering (22). De relatie tussen lotingsklasse en studiesucces (gedefinieerd als het al dan niet halen van de propedeuse in één jaar) is wel duidelijker.

Tabel 19. *Het percentage binnen twee jaar geslaagde studenten bij drie talenstudies*

		Frans	Italiaans	Duits
lotingsklasse	E	16.7	.0	100.0
	D	50.0	27.3	.0
	C	70.0	50.0	.0
	B	50.0	33.3	66.7
	A	100.0	50.0	50.0

Tabel 20. *Het percentage binnen één jaar geslaagde studenten bij drie talenstudies*

		Frans	Italiaans	Duits
lotingsklasse	E	.0	12.5	25.0
	D	.0	13.3	.0
	C	37.5	38.5	.0
	B	33.3	25.0	60.0
	A	75.0	40.0	66.7

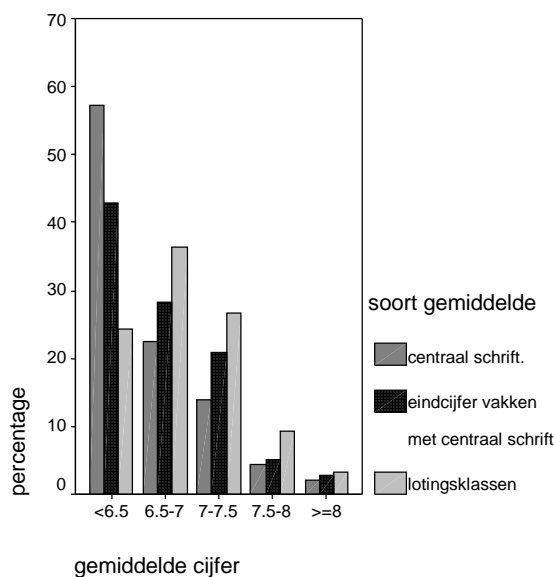
Het cijfer voor VWO-Frans ($N = 57$) loopt van een 6 tot een 9, met een gemiddelde van 7.7. De 7 studenten met een 6 hebben na één jaar de propedeuse niet gehaald. Zes van deze studenten hadden de propedeuse na twee jaar gehaald kunnen hebben; dat blijkt niet te zijn gebeurd. De correlatie tussen het VWO-cijfer Frans en studiesucces is overigens lager dan de correlatie tussen lotingsklasse en studiesucces. Italiaans wordt als VWO-vak op sommige scholen gegeven; het vak kent geen centraal examen. Italiaans als VWO-vak met een cijfer (i.p.v. een waardering) komt in de cohorten die voor Italiaans zijn onderzocht, niet voor. Het VWO-vak Duits is wel door 21 van de 22 studenten Duits gedaan, althans er is van 21 studenten een VWO-cijfer bekend. Het VWO-cijfer loopt van een 6 tot een 9, met een gemiddelde gelijk aan 7.1. de 3 studenten met een 6 hadden hun propedeuse geen van allen na één jaar. Er was slechts één student met een 6 die

minstens twee jaar Duits achter de rug heeft; deze student heeft ook na twee jaar de propedeuse nog niet. Voor Duits is de correlatie tussen de VWO-cijfers en studiesucces hoger dan de correlatie tussen lotingsklasse en studiesucces.

5.6. De verschillende cijfergemiddelden

In de paragrafen 5.1, 5.3 en 5.4 werden bij een deelgroep van de studenten uit de cohort 2003 verschillende cijfergemiddelden onderzocht: het gemiddelde van de cijfers voor het centraal schriftelijk, het gemiddelde eindcijfer voor vakken met een centraal schriftelijk, en het gemiddelde volgens de procedure van de gewogen loting. Het gemiddelde van de eindcijfers voor vakken met een centraal schriftelijk bleek de beste voorspelling op te leveren. In deze paragraaf bekijken wij de overeenkomsten en verschillen tussen de drie soorten gemiddelden voor de totale groep van 811 studenten van wie deze gegevens bekend waren.

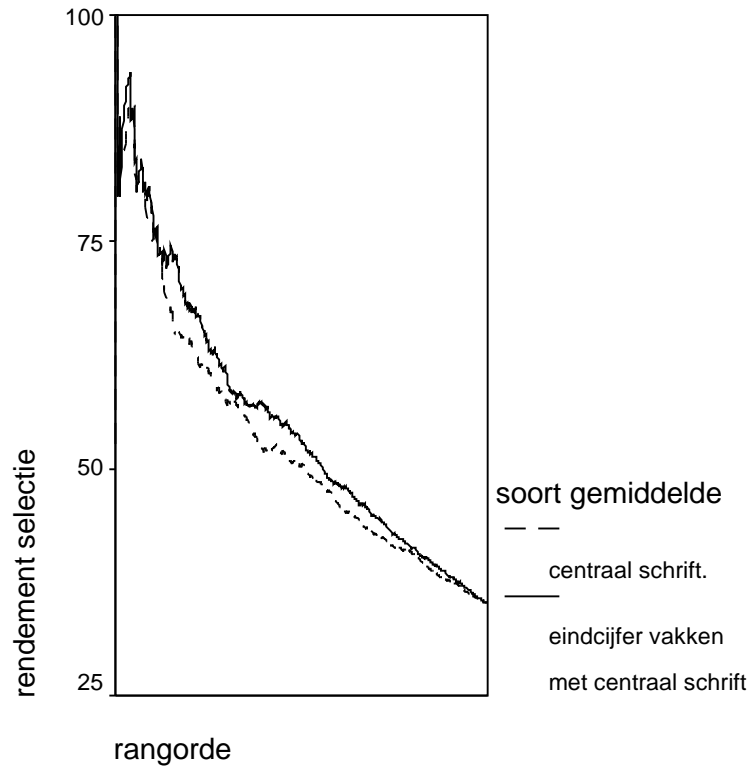
Het gemiddelde van de cijfers voor het centraal schriftelijk en het gemiddelde volgens het systeem van gewogen loting blijken, zoals verwacht, het meest te verschillen. De correlatie tussen deze twee gemiddelden is slechts 0.83; de twee overige correlaties bedragen 0.93. De cijfers volgens het systeem van gewogen loting zijn het hoogst, die voor het centraal schriftelijk het laagst. De verdeling van de cijfers, na een indeling in een aantal cijfercategorieën, staat in Figuur 2. Uit de figuur is op te maken dat het percentage studenten met hoge gemiddelde cijfers (7.5 en hoger, of 8 en hoger) voor de verschillende cijfergemiddelden verschilt.



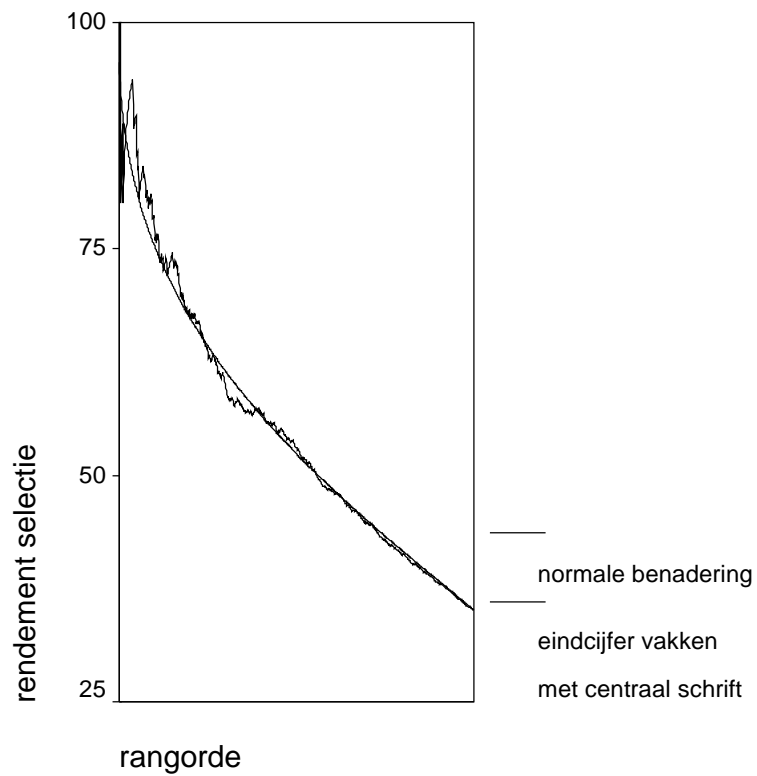
Figuur 2. De verdeling van het gemiddelde cijfer

Het gemiddelde eindcijfer van vakken met een centraal schriftelijk correleert – het is al eerder vermeld – hoger met slagen voor de propedeuse (na één jaar) dan het gemiddelde van de cijfers voor het centraal schriftelijk. Uit de hoogte van de correlatie kan echter niet direct worden afgelezen wat het effect van een eventuele selectie op basis van de gekozen predictor zal zijn. Die vraag trachten wij nu te beantwoorden. Daartoe zijn de studenten naar hun gemiddelde cijfers gerangordend van hoog naar laag (met een willekeurige volgorde binnen een groep studenten met hetzelfde cijfergemiddelde); de student met het hoogste gemiddelde kreeg rang 1, de student met het laagste rang 811. Voor elke rang is het percentage studenten onder de studenten tot en met de desbetreffende rang (de ‘top’ groep) berekend dat geslaagd is voor de propedeuse. Het percentage ‘geslaagd voor de propedeuse’ is vervolgens, in Figuur 3, afgezet tegen het rangnummer. In Figuur 3 staan de uitkomsten voor beide hierboven genoemde berekeningen van het gemiddelde cijfer. Voor beide cijfergemiddelden loopt het rendement omlaag als functie van de rangorde, d.w.z. als functie van het aantal ‘geselecteerde’ studenten. Beide curven vertonen links een enigszins onrustig beeld. Dat is te wijten aan de geringe aantallen waarop de gegevens berusten: als bijvoorbeeld aan de eerste 5 studenten die naar wij aannemen allen zijn geslaagd, een zesde student wordt toegevoegd, die niet geslaagd is, gaat het rendement van de selectie in één sprong omlaag van 100 procent naar 83 procent. In de figuur is te zien dat de curve voor het gemiddelde cijfer gebaseerd op de eindcijfers van de vakken met een centraal schriftelijk over een range van rangordes, d.w.z. over een range van selectieratio’s, boven de curve ligt van het gemiddelde gebaseerd op de cijfers voor het centraal schriftelijk alleen. Het gemiddelde gebaseerd op eindcijfers is dus inderdaad een betere predictor: het levert voor een bepaalde selectieratio een hoger propedeuserendement op. Bij een lage selectieratio zijn beide predictoren echter in termen van rendement vergelijkbaar.

Taylor en Russell (1939) gaven in een reeks bekende tabellen de relatie tussen de effectiviteit van selectie en de hoogte van de validiteitscoëfficiënt onder de veronderstellingen dat aan het dichotome succescriterium een normale verdeling ten grondslag lag en dat de gezamenlijke verdeling van succes en voorspeller bivariaat normaal was. De voorspellers in het huidige onderzoek zijn niet normaal verdeeld. Desalniettemin blijken met de benadering van Taylor en Russell de gegevens uit het huidige onderzoek goed te kunnen beschreven, zo blijkt uit Figuur 4.

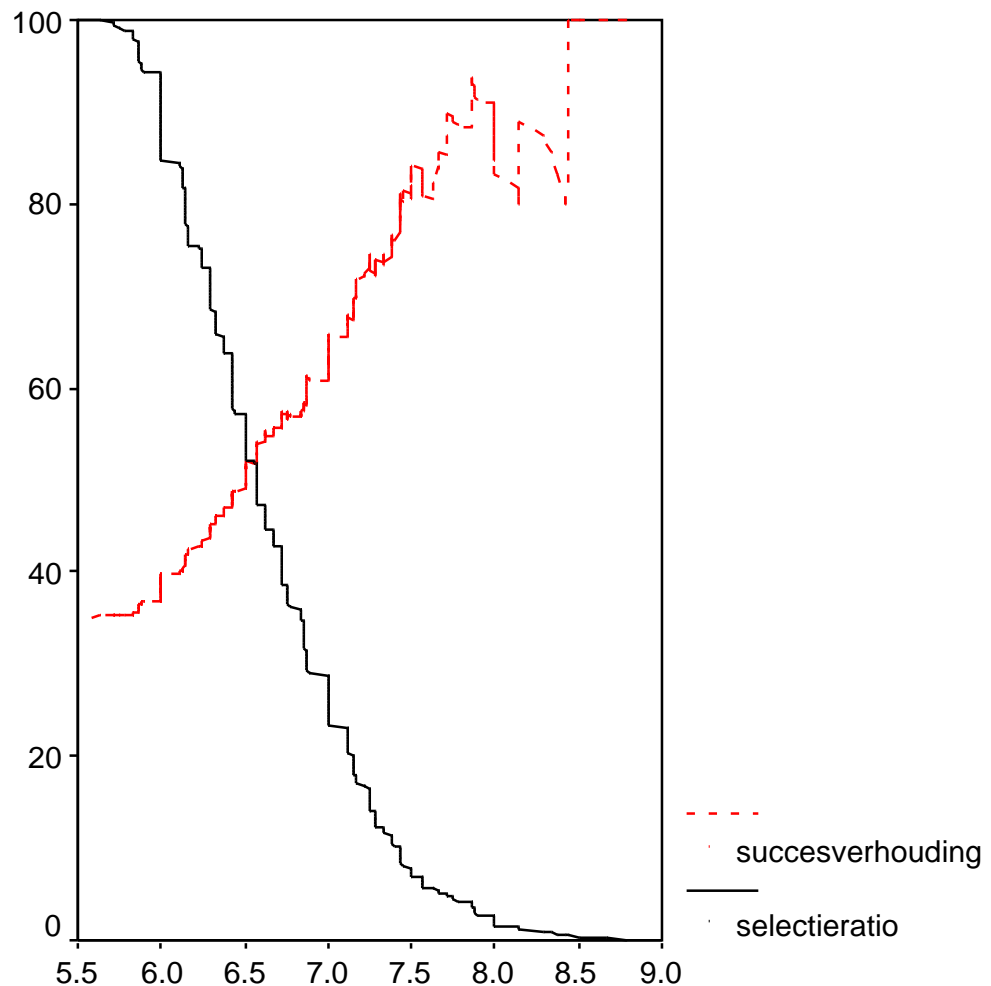


Figuur 3. Het rendement voor twee verschillende predictoren als functie van de selectieratio



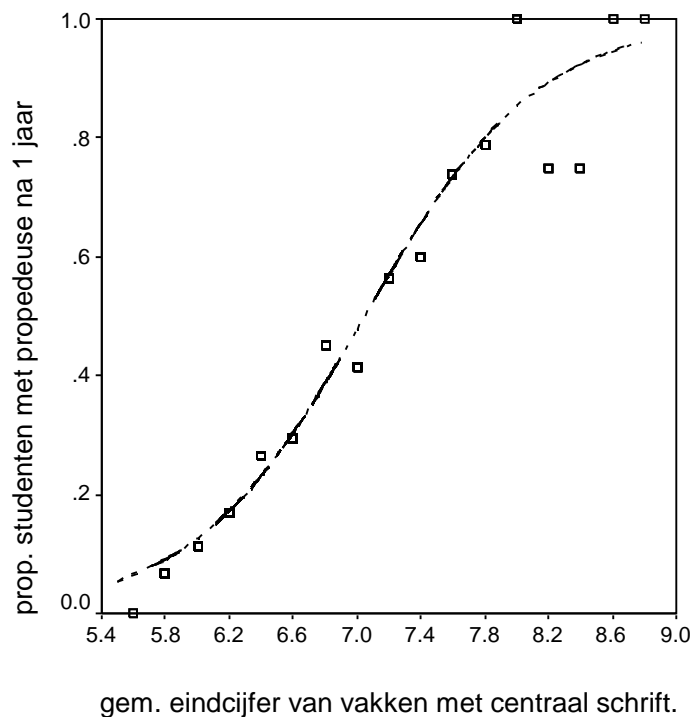
Figuur 4. Het rendement voor het gemiddelde eindcijfer van vakken met een centraal schriftelijk als functie van de selectieratio, en de relatie tussen rendement en predictor gebaseerd op de Taylor & Russell benadering

De resultaten uit deze paragraaf m.b.t. de verdeling van de gemiddelde cijfers en de succesverhouding (percentage ‘voldoende’ kandidaten onder de toegelaten kandidaten) worden samengevat in Figuur 5. In deze figuur zijn de succesverhouding en selectieratio beide als functie van de aftestgrens op de voorspeller *gemiddeld eindcijfer van vakken met een centraal schriftelijk* gegeven. Bij een lage aftestgrens wordt bijna iedereen toegelaten, de selectieratio is hoog. Als men meer naar rechts gaat, ligt de drempel voor toelating hoger en daalt de selectieratio. De succesverhouding ligt relatief laag bij een lage aftestgrens. Naar mate men selectiever is, gaat de succesverhouding omhoog. Uiterst rechts is de empirische relatie tussen de selectieratio en de aftestgrens onregelmatig, zoals dat ook in de Figuren 3 en 4 het geval is. Dat komt omdat het resultaat op een klein aantal gegevens is gebaseerd.



Figuur 5. Selectieratio en succesverhouding als functie van de aftestgrens

Als er daadwerkelijk gaat worden geselecteerd, mag men enig effect verwachten op de samenstelling van de groep kandidaten. Het is aannemelijk dat in die situatie het percentage kandidaten met een relatief laag gemiddeld VWO-cijfer daalt. Dan zullen ook de Figuren 3, 4 en 5 er anders komen uit te zien. Zelfs de kans op succes gegeven het gemiddelde cijfer, weergegeven in Figuur 6, kan veranderen. Dit zal gebeuren als de keuze om al dan niet mee te doen met de selectieprocedure ook binnen de groep kandidaten met hetzelfde cijfergemiddelde samenhangt met studiesucces.



Figuur 6. Succes als functie van het gemiddelde cijfer; via logistische regressie gefitte curve en empirische gegevens (gegroepeerd)

Enige gegevens m.b.t. tot de opbrengst van selectie in verschillende omstandigheden kunnen in de Schema's 1 en 2 gevonden worden.

Schema 1. *Opbrengst van selectie bij ongewijzigde samenstelling van de groep kandidaten*

gem. cijfer	start factor	start aantal	aantal	proportie +	aantal +	aantal selectie	selectie- ratio	aantal succes	succes- ratio
8.5	1	1	1.00	0.92	0.92	1.00	0.01	0.92	0.92
8	1	4	4.00	0.82	3.28	5.00	0.05	4.20	0.84
7.5	1	12	12.00	0.64	7.68	17.00	0.17	11.88	0.70
7	1	22	22.00	0.45	9.9	39.00	0.39	21.78	0.56
6.5	1	36	36.00	0.26	9.36	75.00	0.75	31.14	0.42
6	1	24	24.00	0.14	3.36	99.00	0.99	34.50	0.35
5.5	1	1	1.00	0.08	0.08	100.00	1.00	34.58	0.35
totaal			100.00						

In het schema zijn de gemiddelde cijfers berekend over de vakken met een centraal schriftelijk en afgerond naar hele en halve cijfers. Onder *aantal* staat per cijfercategorie het aantal belangstellenden, het product van *factor* en *start aantal*, rubrieken waar wij later op terug komen. Onder *proportie +* staat de proportie kandidaten in de desbetreffende cijfercategorie die in één jaar de propedeuse haalt; de desbetreffende getallen zijn ontleend aan de feitelijke proporties, maar de relatie tussen gemiddeld cijfer en proportie geslaagd is gladgestreken. Onder *aantal +* staat het aantal geschikte personen in de desbetreffende cijfercategorie: $aantal \times proportie +$. Onder het kopje *aantal selectie* staat het aantal kandidaten dat zou worden toegelaten tot de opleidingen indien de aftestgrens, de cesuur tussen toelating en afwijzing, bij het desbetreffende cijfer zou liggen; *aantal selectie* is de som van de getallen in *aantal +* t/m het desbetreffende cijfer. De *selectieratio* is gelijk aan *aantal selectie* gedeeld door *totaal*. Het *aantal succes* is het aantal succesvolle kandidaten als de cesuur toelating/afwijzing bij het desbetreffende gemiddelde cijfer zou liggen; *aantal succes* is de som van de aantallen onder *aantal +* t/m de desbetreffende cijfercategorie. De *succesratio* is gelijk aan *aantal succes* gedeeld door *aantal selectie*.

In Schema 2 is te zien wat er zou kunnen gebeuren als de omvang en de samenstelling van de groep kandidaten verandert en de succesansen gegeven gemiddeld cijfer (*proportie +*) gelijk blijven. In *factor* staat de factor waarmee het oorspronkelijke aantal kandidaten met een bepaald cijfergemiddelde (*start aantal*) mee moet worden vermenigvuldigd om het aantal kandidaten in de nieuwe situatie te krijgen (*aantal*). Met formules in

een spreadsheet worden de getallen in de overige kolommen direct aangepast.

Schema 2. *Opbrengst van selectie bij gewijzigde samenstelling van de groep kandidaten*

gem. cijfer	start factor	start aantal	aantal	proportie +	aantal +	aantal selectie	selectie-ratio	aantal succes	succes-ratio
8.5	1	1	1.00	0.92	0.92	1.00	0.01	0.92	0.92
8	1	4	4.00	0.82	3.28	5.00	0.07	4.20	0.84
7.5	1	12	12.00	0.64	7.68	17.00	0.23	11.88	0.70
7	1	22	22.00	0.45	9.9	39.00	0.52	21.78	0.56
6.5	1	36	36.00	0.26	9.36	75.00	1.00	31.14	0.42
6	0	24	0.00	0.14	0	75.00	1.00	31.14	0.42
5.5	0	1	0.00	0.08	0	75.00	1.00	31.14	0.42
totaal			75.00						

In Schema 2 is ervan uitgegaan dat zich geen kandidaten meer melden met de laagste cijfergemiddelden (6 en lager). De overige instroom blijft gelijk. Deze situatie kan zich voordoen als bekend is dat de opleidingen alleen kandidaten met een gemiddeld cijfer van 6.5 of hoger accepteren. Het aantal geaccepteerde kandidaten, het aantal succesvolle kandidaten en de succesratio veranderen in deze situatie niet als de cesuur toelating/afwijzing bij 6.5 ligt.

6. De relatie tussen eindexamencijfers en het aantal behaalde studiepunten

Het studiesucces in de eerste fase van de studie kan worden afgemeten aan het al dan niet behalen van de propedeuse binnen een bepaalde termijn. In deze studie is in eerste instantie gekozen voor het al dan niet behalen van de propedeuse binnen twee jaar. Nadeel van dit criterium is dat het studiesucces pas na twee jaar kan worden bepaald. Wij kunnen natuurlijk kijken naar het al dan niet behalen van de propedeuse binnen één jaar, en dat is in deze studie ook gedaan. In dit geval voegen wij de studenten die de propedeuse in de loop van het tweede jaar halen samen met degenen die dat niet doen. Wij beschikken echter over informatie die samenhangt met het al dan niet behalen van de propedeuse in de loop van het tweede jaar: het aantal behaalde studiepunten van het propedeuseprogramma.

Het aantal studiepunten van het propedeuseprogramma is 60 ECTS¹⁶. Een student die 60 ECTS heeft behaald, is voor de propedeuse geslaagd.¹⁷ Bij studenten die de propedeuse niet in het eerste studiejaar hebben gehaald, is het aantal studiepunten een indicatie van het studietempo en een indicatie de kans op slagen in het tweede studiejaar. Over de relatie tussen het aantal studiepunten en de slaagkans kan het volgende worden gezegd. Studenten met minder dan 30 ECTS krijgen aan het eind van het eerste studiejaar een bindend negatief studie-advies (BSA-advies) en mogen niet verder met de studie. Op deze regel zijn uitzonderingen mogelijk: sommige studenten met een onvoldoende aantal studiepunten mogen toch doorgaan wegens persoonlijke omstandigheden of met een beroep op de hardheidsclausule. Bij sommige opleidingen¹⁸ worden extra eisen gesteld en kunnen ook studenten met meer dan 30 ECTS een negatief studieadvies krijgen. Van degenen die met een positief BSA-advies met de studie doorgaan, zijn degenen die veel studiepunten hebben in het voordeel.

¹⁶ Sinds het studiejaar 2002-2003. daarvoor werden STP gehanteerd. Voor de propedeuse stonden 40 STP.

¹⁷ Met minder studiepunten slagen kan niet: indien een compensatieregeling van kracht is, horen studiepunten te worden toegekend aan gecompenseerde onvoldoendes. Het komt wel voor dat een student met 60 of meer studiepunten de propedeuse niet heeft gehaald. Dat kan gebeuren als een student met een major-minor combinatie meer vakken heeft gedaan dan alleen de vakken uit de major-minor combinatie, en de bepaling van het aantal studiepunten niet gebaseerd was op alleen de vakken uit de desbetreffende major-minor combinatie.

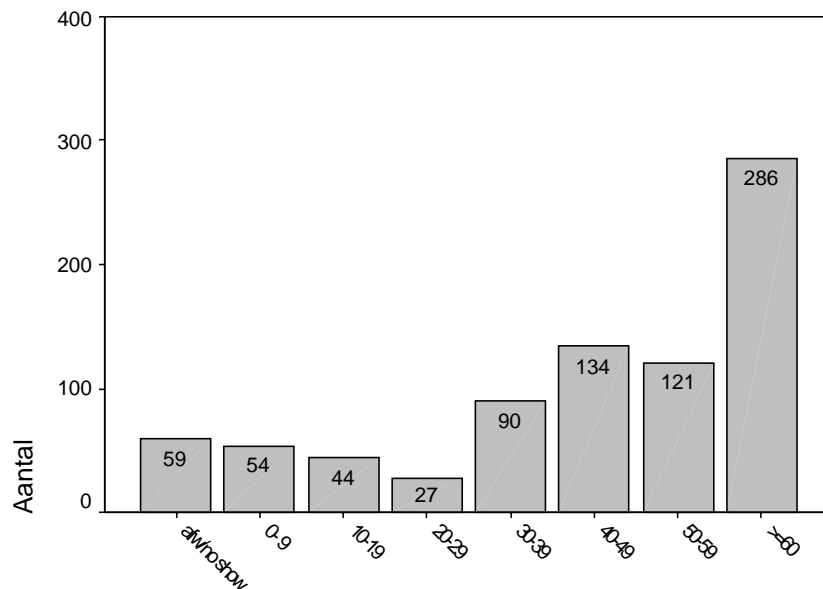
¹⁸ Bij Geschiedenis moet een student bijvoorbeeld voor minstens één van de twee Themacolleges een voldoende hebben.

Het aantal behaalde studiepunten lijkt een goede indicatie van studietempo en daarmee ook voor studiesucces. De relatie tussen studietempo en het aantal behaalde studiepunten is echter gecompliceerder dan het op het eerste gezicht het geval lijkt te zijn. Het is niet zo dat alle studenten met weinig studiepunten een laag studietempo hebben. Een laag aantal studiepunten kan ook een indicatie zijn van het feit dat een student in de loop van het eerste jaar met de studie is gestopt. Enkele studenten, de *no-show* studenten, blijken zelfs nooit een tentamen te hebben gedaan.

De berekening van het aantal behaalde studiepunten is niet eenvoudig. De berekening moet gebaseerd zijn op de voldoende en vrijstellingen die meetellen in het gevolgde propedeuse-programma. In verband met het uitbrengen van het BSA-advies en de twee voortgangsadviezen in het eerste studiejaar is software ontwikkeld. Aan de hand daarvan wordt door de opleidingen het aantal behaalde studiepunten berekend. Het resultaat wordt door de opleidingen naar het centrale niveau doorgegeven. In de praktijk blijken na de waardering in het kader van het BSA van de resultaten door de opleidingen allerlei resultaten niet te kloppen. Dat kan bijvoorbeeld het geval zijn omdat resultaten van late herkansingen alsnog binnen komen of omdat vrijstellingen niet geregistreerd blijken te zijn. De opleidingen evalueren het BSA-advies dan niet opnieuw met het programma dat daarvoor beschikbaar is: de procedure is daarvoor te tijdrovend. De actie die wordt ondernomen hangt af van de aard van de wijzigingen. Indien het BSA-advies positief was, leidt een verhoging van het aantal studiepunten niet tot actie. Als een negatief advies omgezet moet worden naar een positief advies, wordt dat uiteraard wel doorgegeven en verder verwerkt. De consequentie hiervan is dat de voor het onderzoek beschikbare BSA-gegevens niet helemaal correct zijn. Zo blijken er studenten te zijn die voor de propedeuse zijn geslaagd terwijl het aantal geregistreerde studiepunten daarvoor te laag is. Vanwege de fouten in de bestanden werd besloten om alleen de gegevens te analyseren van de studenten van wie de VWO-cijfers in de Access-database zijn opgeslagen. Bij deze groep leken de BSA-resultaten een redelijk beeld te geven. In de gevallen (9) dat het aantal geregistreerde studiepunten minder dan 60 was en de student toch voor de propedeuse was geslaagd, werd het aantal studiepunten op 60 gesteld.

De behaalde studiepunten werden in een aantal klassen verdeeld: 0 – 9, 10 – 19, 20 – 29, 30 - 39, 40 – 49, 50- 59, ≥ 60 en *no-show* (0 studiepunten, en nooit deelgenomen aan een tentamen). Een aantal studenten uit de onderzochte groep

bleek geen BSA-advies te hebben gekregen; dat kan omdat studenten die zich tijdig afmelden niet onder het BSA-regiem van het desbetreffende cursusjaar vallen. In Figuur 7 staat de verdeling van de BSA-adviezen, waarbij de categorieën *no-show* (9 studenten) en ontbrekend advies zijn samengenomen.



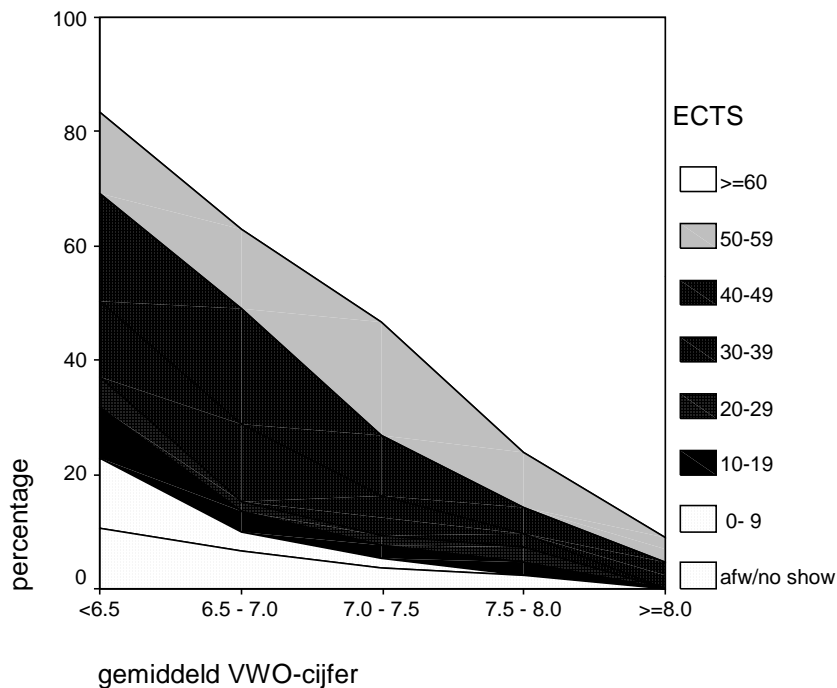
ECTS

Figuur 7. De verdeling van het aantal behaalde studiepunten

De verdeling is scheef: de meeste studenten zitten aan de rechterkant en de grootste categorie is de categorie ≥ 60 . Praktisch alle studenten uit deze groep hebben de propedeuse behaald. Links is er een relatief grote groep studenten die in feite niet heeft meegedaan: de studenten die geen advies hebben gekregen en de studenten die als *no-show* studenten zijn geregistreerd. De categorie 20-29 ECTS is relatief klein, zeker in vergelijking met de categorie 30-39 ECTS. Dit is een gevolg van het BSA-systeem dat studenten motiveert om in ieder geval 30 ECTS te behalen omdat zij anders het recht op verdere studie bij de opleiding verliezen.

De relatie tussen VWO-cijfers en het aantal behaalde studiepunten wordt weergegeven in Figuur 8. Bij de VWO-cijfers is het gemiddelde cijfer op de eindexamenvakken die ook een centraal gedeelte kennen, gebruikt. De percentages voor de variabele *aantal behaalde studiepunten* tellen binnen elke klasse van VWO-cijfers op tot 100 procent. In de figuur is fraai te zien dat niet alleen het percentage studenten dat de propedeuse in één jaar (≥ 60 ECTS) haalt, geleidelijk

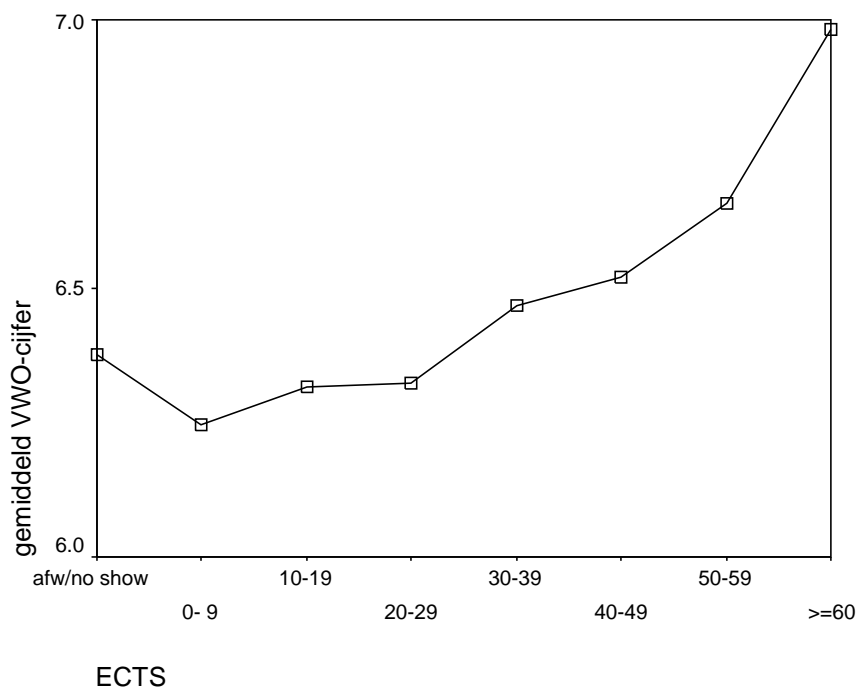
toeneemt met het gemiddeld VWO-cijfer. Ook de percentages studenten met een groot aantal studiepunten, bijvoorbeeld 30 of meer, loopt op. Het percentage studenten dat uitvalt, d.w.z. geen BSA-advies heeft gekregen, of geen enkel cijfer heeft behaald, neemt af met het gemiddeld VWO-cijfer. In de hoogste categorie VWO-cijfers blijkt bijna iedereen de propedeuse (≥ 60 ECTS) te halen en voor zover dat niet gebeurt, veel studiepunten. Daarentegen blijken er in de laagste categorie VWO-cijfers studenten voor te komen die in één jaar de propedeuse halen.



Figuur 8. De relatie tussen VWO-cijfers en het aantal behaalde studiepunten.

Wij kunnen de relatie tussen VWO-cijfers en het aantal behaalde studiepunten ook op een andere manier bekijken. In Figuur 9 wordt voor elke categorie van de variabele *studiepunten* het gemiddelde eindexamencijfer, berekend voor de vakken met een centraal deel, gegeven. De groep studenten met 60 of meer ECTS springt er qua VWO-cijfers uit. Uit de figuur blijkt ook dat de gemiddelde VWO-prestatie van de studenten in de lagere categorieën niet toeneemt. Dat is niet zo vreemd. Zoals reeds is opgemerkt is het aantal behaalde studiepunten geen eenvoudige maat van studietempo: een laag aantal behaalde studiepunten geeft aan dat een student weinig vakken heeft gehaald of dat een student vroegtijdig is gestopt, of een combinatie van beide. Dit gegeven en het feit dat minder dan 30 studiepunten in principe tot een negatief BSA leidt, suggereert dat wij voor een verdere analyse van

de relatie tussen VWO-cijfers en behaalde studiepunten de categorieën tot 30 studiepunten moeten samennemen. Wij krijgen dan de volgende ‘kwantificatie’ van de categorieën: 1 ‘minder dan 30 ECTS’, 2 ‘30-39 ECTS’, 3 ‘40-49 ECTS’ 4 ‘50-59 ECTS’ 5 ‘ ≥ 60 ’¹⁹. De correlatie tussen deze scoring van het aantal behaalde studiepunten en VWO-cijfers (gemiddelde voor de vakken met een centraal deel) is 0.46, nauwelijks hoger dan de correlatie tussen het al dan niet in één keer behalen van de propedeuse en VWO-cijfers, 0.43. Wij moeten natuurlijk ook naar de correlaties voor de afzonderlijke opleidingen kijken. Het beeld dat wij dan krijgen, is wisselend. Bij Rechten is de correlatie met VWO-cijfers zelfs iets lager dan de correlatie tussen het al dan niet behalen van de propedeuse met VWO-cijfers, voor Psychologie en Geschiedenis is de correlatie hoger. Bij Geschiedenis is het verschil tussen de correlaties het grootst: de correlatie tussen de gekwantificeerde studiepunten-categorieën en VWO-cijfers is 0.63, de correlatie tussen het al dan niet behalen van de propedeuse en VWO-cijfers is 0.52.

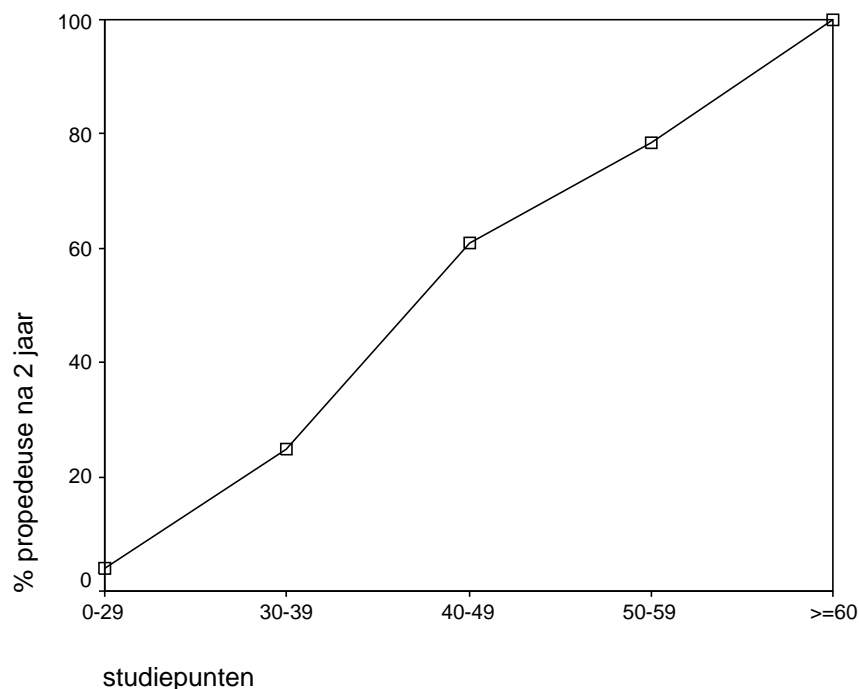


Figuur 9. Het gemiddelde eindexamencijfer per categorie ECTS.

Hoe nuttig is het nu om het aantal behaalde studiepunten als criterium te gebruiken naast het al dan niet behalen van de propedeuse in één, resp. twee jaar? In de eerste

¹⁹ Als wij de correlatie tussen VWO-cijfer en de studiepunten-categorieën zouden willen maximaliseren, dan zouden wij aan elke studiepunten-categorie het gemiddelde VWO-cijfer moeten toekennen van de studenten uit de desbetreffende categorie. De score van de categorie ‘ ≥ 60 ’ zou in dat geval relatief hoger uitvallen, zo kan worden afgelezen uit Figuur 9.

alinea is daar iets over gezegd: het aantal behaalde studiepunten geeft na één jaar informatie over de variabele *al dan geen propedeuse in twee jaar*, waarover pas na twee jaar kan worden beschikt. Meer concrete informatie over de waarde van het aantal behaalde studiepunten kan worden gegeven indien ook het propedeuse-succes na twee jaar bekend is. Gegevens hierover zijn beschikbaar, bijvoorbeeld voor de cohort 2002. Deze gegevens hebben wij tot nu toe wegens de onbetrouwbaarheid ervan, niet gebruikt. Voor de beantwoording van de vraag naar het nut van de studiepunten variabele zullen wij echter alsnog van deze gegevens gebruik maken. Wij beschikten over de gegevens van 1097 studenten uit de cohorten 2002 Duits, Bestuurskunde, Frans, Geschiedenis, Italiaans, Psychologie en Rechten. De relatie tussen het aantal behaalde studiepunten en al dan niet binnen twee jaar slagen voor de propedeuse staat in Figuur 10²⁰.



Figuur 10. Relatie tussen in het eerste jaar behaalde studiepunten en al dan geen propedeuse na twee jaar.

Alle studenten in de studiepunten-categorie ≥ 60 hebben na twee jaar de propedeuse. In de groep studenten met minder dan 30 studiepunten zitten enkele studenten die geen negatief bindend studieadvies hebben gekregen en de propedeuse alsnog binnen twee jaar hebben behaald. De kans op succes loopt sterk op van de categorie *30-39 studiepunten* tot de categorie *50-59 studiepunten*. Het

²⁰ Studenten zonder BSA-advies zijn aan de categorie *0-29 studiepunten* toegewezen.

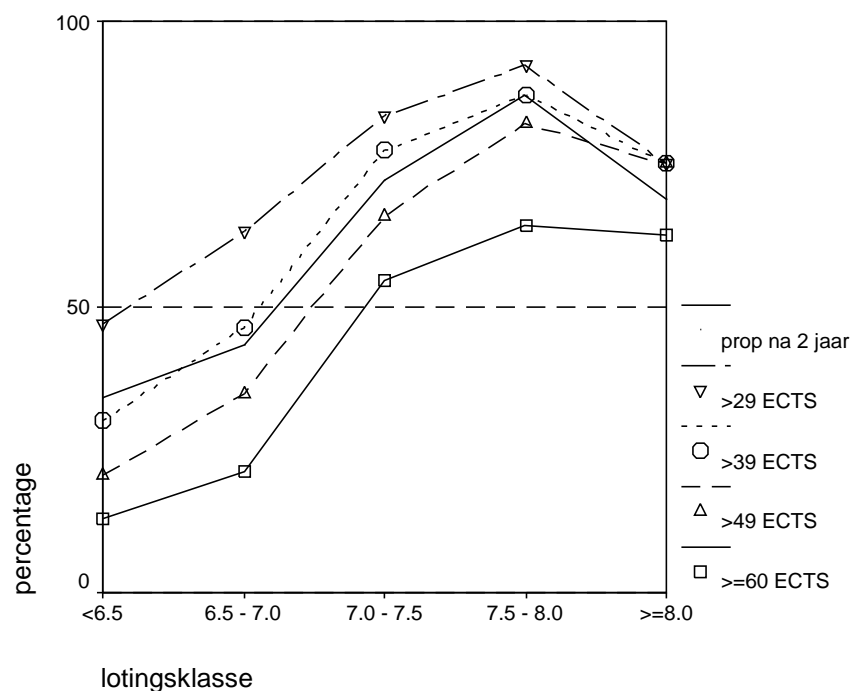
studiepunten criterium lijkt inderdaad een goede benadering na één jaar te geven van het al dan niet na twee jaar behaald hebben van de propedeuse. De correlatie tussen het propedeuseresultaat na twee jaar en de studiepunten ingedeeld in categorieën zoals in Figuur 10 (met opeenvolgende integers als scores voor de categorieën), is 0.80, hoger dan de correlatie tussen het propedeuseresultaat na één jaar en dat na twee jaar ($r = 0.63$).

Wij kunnen natuurlijk in plaats van twee verschillende criteria gebaseerd op het al dan niet halen van de propedeuse, de stand van zaken na één jaar en die na twee jaar, een criterium construeren waarbij de datum van het behalen van de propedeuse een rol speelt. Wij zouden bijvoorbeeld het behalen van de propedeuse in het tweede jaar een lagere waardering kunnen meegeven dan het behalen van de propedeuse in het eerste jaar. Aan de hand van de voorspeller, het gemiddeld eindexamencijfer (gebaseerd op een indeling in 5 lotingsklassen met scores 1 voor de categorie E t/m 5 voor de categorie A) verschilt de groep studenten die de propedeuse in het eerste studiejaar haalt, sterk van de groep die de propedeuse in het tweede jaar haalt²¹. De gemiddelde eindexamencijfers lopen uiteen van 1.8 voor de groep studenten die de propedeuse niet haalt, en 2.0 voor de groep die de propedeuse in het tweede jaar haalt tot 2.8 voor de groep die de propedeuse in het eerste jaar haalt. Qua gemiddelde VWO-cijfer lijkt de groep studenten die in het tweede jaar de propedeuse haalt, meer op de groep die de propedeuse niet haalt dan op de groep die de propedeuse in het eerste jaar haalt. Als wij het behalen van de propedeuse in het tweede jaar met een factor een half wegen, is de correlatie tussen het aantal behaalde studiepunten en al dan niet slagen voor de propedeuse gelijk aan 0.89. Hiermee is het aantal behaalde studiepunten, met de hier gebruikte indeling in categorieën, een nuttig criterium ter aanvulling op of vervanging van het criterium *al dan niet geslaagd binnen twee jaar*. Indien het aantal studiepunten ook in de toekomst in selectie-onderzoek gebruikt wordt, is het wel zaak er voor te zorgen dat het aantal studiepunten beter dan nu wordt vastgelegd.

Bij de deelgroepen uit de cohort 2002 met studenten Rechten, Psychologie en Geschiedenis bestuderen wij de relatie tussen VWO-cijfers en het aantal behaalde studiepunten, weergegeven in Figuur 8, opnieuw. De relatie tussen beide variabelen wordt op iets andere wijze weergegeven, in Figuur 11 voor Rechten, in Figuur 12 voor Psychologie, en in Figuur 13 voor Geschiedenis. In Figuur 11 is af

²¹ De groepen verschillen ook in gemiddeld cijfer voor eerstejaarsvakken. Meer informatie daarover staat in Bijlage 1.

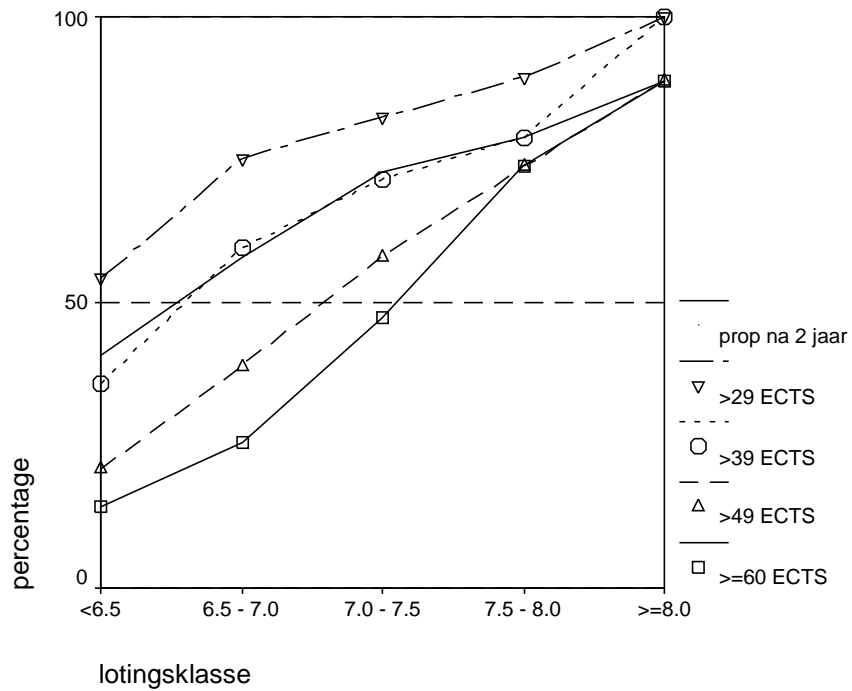
te lezen hoe het percentage studenten met de waardering ≥ 60 ECTS (praktisch *het behalen van de propedeuse in 1 jaar*) toeneemt met het gemiddelde VWO-cijfer. Ook kan worden afgelezen hoe de percentages toenemen voor > 49 ECTS, > 39 ECTS en > 29 ECTS. Bovendien is in de figuur het percentage studenten dat na twee jaar voor de propedeuse geslaagd is, weergegeven als functie van het gemiddeld VWO-cijfer. Bij *Rechten* wijkt de curve voor > 39 ECTS niet erg af van de curve voor het al dan niet behalen van de propedeuse in twee jaar. Voor beide geldt dat het percentage bij de categorie VWO-cijfer 6.5 – 7.0 net onder de 50 procent ligt en bij de categorie 7.0 – 7.5 ruim boven de 50 procent. De percentages dalen bij de hoogste categorie VWO-cijfers. Dit heeft te maken met steekproeffluctuaties. De gegevens voor de hoogste lotingsklasse zijn gebaseerd op een gering aantal studenten. Met name hier blijkt dat de empirische curves voor daadwerkelijk gebruik niet voldoen en gladgestreken zouden moeten worden.



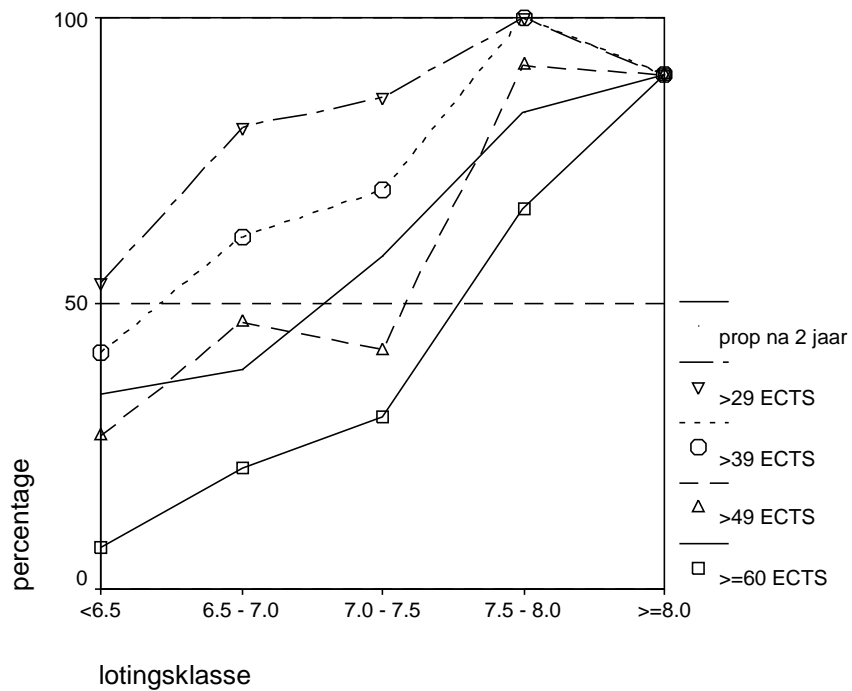
Figuur 11. Relatie tussen VWO-cijfers en behaalde studiepunten: *Rechten*.

Bij Psychologie ligt het percentage studenten met meer dan 39 studiepunten (en propedeusediploma) al bij categorie 6.5 – 7.0 boven de 50 procent. Bij Geschiedenis is het percentage studenten met meer dan 39 studiepunten ook al groter dan 50 procent in de categorie 6.5 – 7.0. Bij Geschiedenis blijft in deze klasse van VWO-cijfers het percentage binnen twee jaar geslaagde studenten echter achter. Bij deze opleiding lijken de curves voor het al dan niet binnen twee jaar

behalen van de propedeuse en die voor meer dan 49 studiepunten meer op elkaar. Gezien de onnauwkeurigheid rond de bepaling van het aantal behaalde studiepunten, moet een definitief oordeel over verschillen tussen opleidingen echter achterwege blijven.



Figuur 12. Relatie tussen VWO-cijfers en behaalde studiepunten: Psychologie.



Figuur 13. Relatie tussen VWO-cijfers en behaalde studiepunten: Geschiedenis.

7. Het vergelijkbaar maken van cijfers op heelvakken en deelvakken

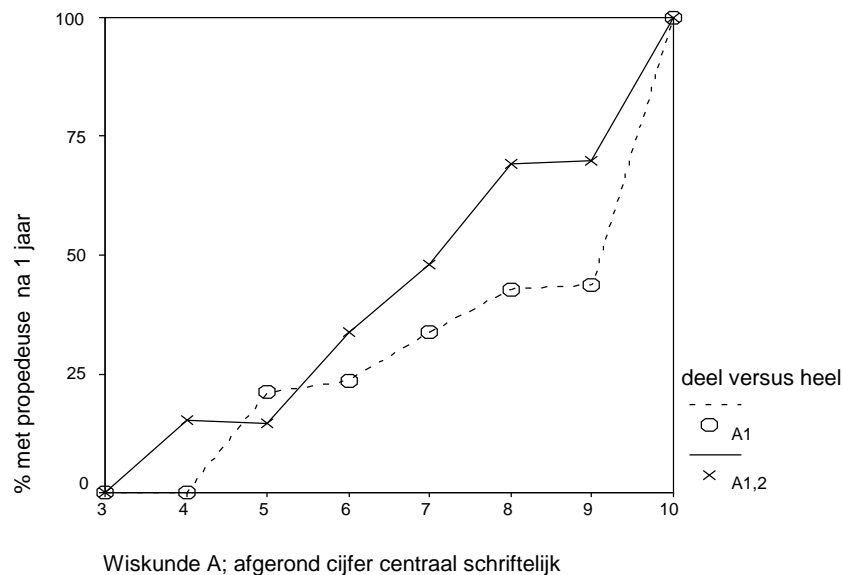
In de voorafgaande hoofdstukken is er meerdere keren op gewezen dat de gebruikte gemiddelde cijfers voor verschillende studenten gebaseerd zijn op cijfers voor verschillende vakken. De cijfers voor het ene vak, gedaan door een deelgroep van studenten A, kunnen hoger uitvallen dan de cijfers voor een ander vak, gedaan door een andere groep, groep B. Zo blijkt zowel het gemiddelde eindcijfer voor Wiskunde A1 als het gemiddelde voor het centraal schriftelijk examen Wiskunde A1 hoger dan het gemiddelde eindcijfer en centraal schriftelijk voor Wiskunde A1,2 (zie Tabel 1). De vraag is wat de oorzaak van deze verschillen is. Is de groep studenten die Wiskunde A1 heeft gedaan en zich in 2003 voor één van de studies Psychologie, Geschiedenis of Rechten (op deze subgroep zijn de gegevens uit Tabel 1 gebaseerd) heeft aangemeld, beter dan de groep die het examen Wiskunde A1,2 heeft gedaan, of is het examen Wiskunde A1,2 moeilijker en stelt bijvoorbeeld een 7 voor Wiskunde A1,2 meer voor dan een 7 voor Wiskunde A1? De meer algemene vraag is die naar de vergelijkbaarheid van cijfers voor verschillende vakken.

De vraag naar de vergelijkbaarheid van de cijfers Wiskunde A1 en Wiskunde A1,2 is een voor de hand liggende vraag, een die meer voor de hand ligt dan de vraag naar de vergelijkbaarheid van zeg het cijfer voor Geschiedenis en Natuurkunde 1.²² Immers, Wiskunde A1,2 is een uitbreiding van de stof van Wiskunde A1: Wiskunde A1 is het deelvak Wiskunde A, en Wiskunde A1,2 is het heelvak Wiskunde A. De centraal schriftelijke examens Wiskunde A1 en Wiskunde A1,2 eerste tijdvak van een bepaald examenjaar hebben dan ook een aantal vragen gemeenschappelijk. Onder de veronderstelling dat de tests dezelfde vaardigheid meten, een veronderstelling waar wij nog op terug komen, is het mogelijk Wiskunde A1 en Wiskunde A1,2 te equivaleren²³, d.w.z. de scores op Wiskunde A1 en Wiskunde A1,2 naar een gemeenschappelijke schaal te transformeren, en daarmee ook de cijfers van Wiskunde A1,2 te vertalen naar de cijferschaal van Wiskunde A1. Voordat wij die optie nader bestuderen, onderzoeken wij de mogelijkheid om vergelijkbare scores te verkrijgen zonder een beroep te doen op de veronderstelling dat de examens op een eendimensionale schaal liggen.

²² Voor verschillen tussen dergelijke vakken kan wellicht gedeeltelijk worden gecompenseerd door opname van profielkeuze als predictor. Ook een techniek voor ontbrekende data (Smits e.a., 2002) biedt perspectief als de gegevens dat toelaten.

²³ Een tweede eis voor het verkrijgen van equivalente scores, d.w.z. scores met dezelfde precisie, is dat de meetinstrumenten even nauwkeurig meten.

Angoff (1991) geeft verschillende definities van ‘vergelijkbare’ scores. Eén van deze opties ligt in het huidige onderzoek voor de hand: een score op X_1 is vergelijkbaar met een score op X_2 indien beide scores dezelfde waarde op criterium Y voorspellen. Het gaat in ons onderzoek immers om een optimale voorspelbaarheid van een succescriterium. Uiteraard is met bovenstaande definitie de vergelijkbaarheid van scores niet alleen afhankelijk van de samenstelling van de groepen personen die een score op X_1 , resp. X_2 hebben behaald, maar ook van de keuze van het criterium. Hier wordt het al dan niet binnen één jaar behalen van de propedeuse als criterium gehanteerd, X_1 is het cijfer op het centraal schriftelijk voor Wiskunde A1, en X_2 is het cijfer voor het centraal schriftelijk voor Wiskunde A1,2. De regressie van het criterium op de cijfers voor Wiskunde A1 en Wiskunde A1,2 is grafisch weergegeven in Figuur 14.



Figuur 14. De regressie van het al dan niet behaald hebben van de propedeuse op het cijfer voor Wiskunde A1, resp. Wiskunde A1,2.

Ter vereenvoudiging benaderen wij de regressie van Y op X_1 met een rechte, evenals de regressie van Y op X_2 .²⁴ Voorts veronderstellen wij dat de twee regressielijnen dezelfde helling hebben. Het verschil tussen de schalen van X_1 en X_2 kan met een additieve constante C_2 weergegeven worden. De waarde van deze constante is op te lossen uit de volgende regressievergelijking:

²⁴ Vanwege het dichotome criterium ligt een logistische regressie voor de hand (zie Van Houwelingen, Stijn, & van Strik, 1995). Wij kiezen hier voor de eenvoud. In de regressievergelijking nemen wij evenmin een code voor de opleidingen op.

$$\hat{Y} = b_0 + b_1(X_1 + X_2 + C_2 X_3), \text{ waarbij } X_3 = 0 \text{ voor } X_1 > 0; X_3 = 1 \text{ voor } X_2 > 0 \\ = b_0 + b_1(X_1 + X_2) + b_3 X_3.$$

Aangezien een student of een cijfer voor X_1 (Wiskunde A1) heeft, of een cijfer voor X_2 (Wiskunde A1,2) is de som $X_1 + X_2$ gelijk aan het behaalde cijfer. In bovenstaande formule is X_3 een indicator die aangeeft of een student een cijfer op X_1 heeft behaald of op X_2 . De regressieanalyse²⁵ levert een schatting op $C_2 = 0.96$. Dat wil zeggen, om de cijfers voor Wiskunde A1 en Wiskunde A1,2 ruwweg vergelijkbaar te maken, moeten wij eerst 0.96 bij het cijfer voor Wiskunde A1,2 optellen.

Wij kunnen het resultaat uit de regressieanalyse vergelijken met de uitkomst van de analyse die gebaseerd is op alleen de gegevens van de centrale examens voor Wiskunde A1 en wiskunde A1,2. Daarbij past een kanttekening. Bij de regressieanalyse, waarbij de cijfers via een extern criterium vergeleken worden, weerspiegelt de additieve constante niet noodzakelijkerwijze alleen het verschil in zwaarte van Wiskunde A1 en Wiskunde A1,2. De groep studenten die Wiskunde A1,2 heeft gedaan i.p.v. Wiskunde A1, verschilt niet alleen daarin van de groep die Wiskunde A1 heeft gedaan. Wiskunde A1,2 is onderdeel van het profiel *E&M*. Leerlingen met het profiel *C&M* kunnen daarentegen volstaan met het deelvak Wiskunde A1. Een mogelijk verschil in cijferschaal voor het heelvak en het deelvak is statistisch gezien verstrengeld met het onderscheid tussen twee profielen; dat zou kunnen worden onderzocht door het profiel als voorspeller te hanteren. Aangezien de verschillende opleidingen in het onderzoek - Rechten, Geschiedenis en Psychologie – een verschillende aantrekkingskracht hebben voor studenten met verschillende profielen, kan een deel van de omvang van het effect C_2 ook met de interactie tussen profielkeuze en keuze voor vervolgopleiding te maken hebben.

Een ander probleem met bovenstaande benadering van het vraagstuk van vergelijkbare scores heeft te maken met het feit dat de relatie tussen X_1 en X_2 in de loop van de jaren kan veranderen. Gegevens over de verandering, bijvoorbeeld een verandering in de additieve constante C_2 , zijn echter pas beschikbaar als de

²⁵ Enkele studenten kwamen twee keer voor; zie voetnoot 12. Bij deze studenten is in de analyse alleen het eerste record meegenomen.

criteriumscores bekend zijn, en dat is te laat als de scores X_1 en X_2 gebruikt moeten worden voor selectiedoeleinden²⁶.

Bovenstaande complicaties doen zich niet voor bij het interne equivaleringsonderzoek m.b.t. de prestaties op de centrale toetsen voor Wiskunde A1 en Wiskunde A1,2, onder de veronderstelling dat de antwoorden op de vragen uit de examens redelijk te beschrijven zijn met een ééndimensionaal antwoordmodel²⁷. Bij het interne equivaleringsonderzoek wordt gebruik gemaakt van het feit dat de examens voor Wiskunde A1 en Wiskunde A1,2 items gemeenschappelijk hebben. Via deze gemeenschappelijke items kunnen de beide examens op een gemeenschappelijke schaal gebracht worden. Als bijvoorbeeld zou blijken dat leerlingen die het Wiskunde A1,2 examen hebben afgelegd, gemiddeld een lager cijfer halen dan – op de gemeenschappelijke items – vergelijkbare leerlingen die Wiskunde A1 hebben gedaan, dan zou dat een indicatie zijn dat de becijfering van het Wiskunde A1,2 examen relatief te zwaar was. Het equivaleringsonderzoek leidt tot een omzetting van de cijfers van Wiskunde A1,2 naar de cijferschaal van Wiskunde A1 (of omgekeerd), zodanig dat de cijfers na de omzetting equivalent zijn.²⁸

In een CITO-onderzoek met landelijke gegevens (zie Bijlage 2) is de relatie tussen heelvak-deelvak onderzocht voor de jaren 2002, 2003 en 2004. Eén van de uitkomsten van dit onderzoek is dat er in 2002 en 2003 gemiddeld 1 cijferpunt extra gerekend zou mogen worden voor Wiskunde A1,2. Voor 2004 zou gemiddeld 0.9 cijferpunt extra gegeven moeten worden om Wiskunde A1,2 op met Wiskunde A1 vergelijkbare wijze te waarderen. Deze gemiddelde cijferaanpassing correspondeert heel goed met de uitkomst van de eenvoudige regressieanalyse waarmee wij zijn begonnen. Alvorens verder te gaan, moeten wij stilstaan bij een belangrijke uitkomst van het CITO-onderzoek, namelijk het feit dat de verschillen ruwweg gelijk zijn van jaar tot jaar. Het lijkt onwaarschijnlijk dat de becijfering van Wiskunde A1 van jaar tot jaar verschilt, evenals die van Wiskunde A1,2, maar dat de relatie tussen Wiskunde A1 en Wiskunde A1,2 relatief constant is. Over de jaren heen stabiele resultaten voor Wiskunde A1, en Wiskunde A1,2 zijn van groot

²⁶ Wij kunnen i.p.v. het criterium al dan niet binnen één jaar geslaagd voor de propedeuse de score op één van de algemene vakken Nederlands of Engels nemen.

²⁷ Uiteraard wordt hiermee niet betoogd dat leerlingen die Wiskunde A1 hebben gedaan, vragen uit de door hen niet bestudeerde Wiskunde A1,2 stof op dezelfde wijze kunnen beantwoorden als leerlingen die deze stof wel hebben bestudeerd.

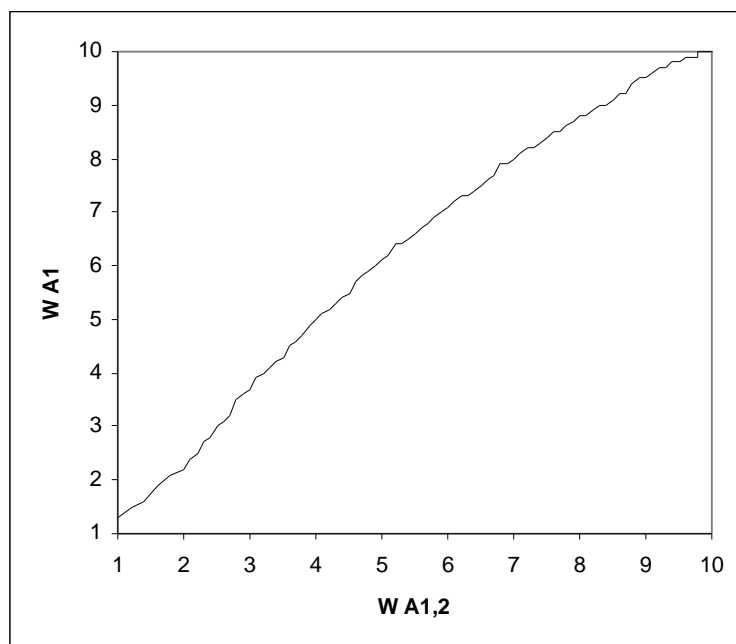
²⁸ Deelvak en heelvak liggen op aparte schalen omdat zij gedefinieerd zijn als verschillende vakken met verschillende eindtermen.

belang. Stel voor dat de resultaten voor beide onderdelen teveel zouden verschillen, zoals de resultaten voor eindexamenvakken in het verleden van jaar tot jaar varieerden (Van den Bergh *et al.*, 2003). Dan zouden wij met een omzetting van het cijfer voor Wiskunde A1,2 naar de schaal van Wiskunde A1 weliswaar bereiken dat de eindexamencijfers binnen één jaar vergelijkbaar zouden zijn, maar wij zouden tevens het risico lopen dat de verschillen tussen jaren soms zouden toenemen i.p.v. afnemen. Indien bijvoorbeeld in een bepaald jaar de becijfering voor Wiskunde A1 te hoog zou zijn uitgevallen, dan zou met de correctie van Wiskunde A1,2 cijfers ook Wiskunde A1,2 voor dat jaar te hoog uitvallen.

Een tweede belangrijke conclusie die wij uit de stabiliteit van de Wiskunde A cijfers mogen trekken, is dat het niet al te veel zal uitmaken of men bij de correctie van de cijfers uit 2002 de omzetting van Wiskunde A1,2 naar de schaal van Wiskunde A1 uit het jaar 2002 gebruikt, of dat men de oude omzettingstabel uit 2003 hanteert. Wij zullen daarom bij de deelgroep van de instroom 2003 van wie wij centrale examencijfers hebben, voor de omzetting van de Wiskunde A1,2 cijfers naar de schaal van Wiskunde A1 voor diploma's uit alle jaren de door het CITO berekende omzettingstabel voor 2003 hanteren^{29,30}. In Figuur 15 is de relatie tussen de cijfers voor Wiskunde A1,2 en Wiskunde A1 gegeven, voor de situatie waarin het cijfer Wiskunde A1,2 omgezet moet worden naar de cijferschaal van Wiskunde A1.

²⁹ 85 procent van de 811 studenten uit dit onderzoek heeft het VWO-diploma in 2003 gehaald.

³⁰ In 2003 ontbreken er enkele cijfers die in de omzettingstabel die in andere jaren wel kunnen voorkomen. Daarvoor is in de omzettingstabel geïnterpoleerd.



Figuur 15. Relatie tussen Wiskunde A1,2 en Wiskunde A1, t.b.v. de omzetting van Wiskunde A1,2 cijfers; gebaseerd op het examen 2003

Bij Wiskunde B is het verschil tussen het centraal schriftelijk voor het deelvak en het centraal schriftelijk voor het heelvak kleiner. In 2002 zou voor Wiskunde B1,2 gemiddeld 0.7 punt bij het cijfer moeten worden opgeteld om het cijfer vergelijkbaar te maken met het cijfer voor Wiskunde B1, in 2003 gaat het om 0.6 punt en in 2004 (niet in dit onderzoek betrokken) om 0.5 punt. De relatie tussen Wiskunde B1,2 en Wiskunde B1 varieert van jaar tot jaar en het verschil tussen de becijferingen lijkt kleiner te worden. Het lijkt verantwoord om in het huidige onderzoek voor alle diploma's de omzettingstabel uit 2003 gebruiken.

Na omzetting van de cijfers voor Wiskunde A1,2 naar de cijferschaal van Wiskunde A1, en die van Wiskunde B1,2 naar de cijferschaal van Wiskunde B1, rijst de vraag hoe de cijferschalen voor Wiskunde A en die voor Wiskunde B kunnen worden gecombineerd tot één W(iskunde)-schaal met voor elke student één W-cijfer³¹. Daarvoor hanteren wij de regressietechniek die aan het begin van dit hoofdstuk is toegepast bij de eerste vergelijking tussen Wiskunde A1,2 en Wiskunde A1. Het blijkt dat er een additieve constante gelijk aan 0.84 bij de

³¹ Alle 811 studenten uit het onderzoek hebben een cijfer voor Wiskunde. Drie studenten hadden zowel een cijfer voor A als voor B. Bij deze studenten is het hoogste van de twee cijfers in de verdere berekeningen m.b.t. de W-schaal meegenomen.

Wiskunde B cijfers moet worden opgeteld om de cijfers voor Wiskunde A en Wiskunde B op een gemeenschappelijke W-schaal te krijgen³².

Vervolgens werden de correlaties berekend tussen P , al dan niet geslaagd voor de propedeuse, en:

W : het 'cijfer' op de W-schaal,

$W(CO)$: het ongecorrigeerde cijfer voor het centraal schriftelijk; bij twee cijfers is het hoogste genomen

$W(EO)$: het eindcijfer; bij twee cijfers is het hoogste genomen

$W(EC)$: het hoogste gecorrigeerde eindcijfer; voor de correcties werd de helft van de additieve correcties gebruikt die bij het onderzoek naar de cijfers voor het centraal schriftelijk werden gevonden.

De correlaties binnen de totale groep van 811 studenten waren:

r_{PW} : 0.30

$r_{PW(CO)}$: 0.29

$r_{PW(EO)}$: 0.34

$r_{PW(EC)}$: 0.35.

Het blijkt dus dat de schaalcorrectie niet in een verbetering van de samenhang tussen wiskunde-resultaat en succes in de vervolgopleiding resulteert. Het eindexamencijfer, inclusief het resultaat voor het schoolonderzoek, levert de beste voorspelling op.

Er zijn drie andere vakken waarbij onderscheid gemaakt wordt tussen een deelvak en een heelvak (1,2 versus 1), te weten Economie, Natuurkunde en Scheikunde. Bij geen van deze vakken zullen wij de analyse herhalen zoals wij die voor wiskunde hebben gedaan. Eén argument is dat zelfs als wij heelvak en deelvak combineren om tot één cijfer te komen, slechts een deel van de studenten een resultaat heeft. Natuurkunde en Scheikunde trekken nog geen 200 van de 811 studenten (zie Tabel 1). Economie is als Economie 1,2 of als Economie 1 door een grote groep studenten gedaan, maar bij Economie hebben het heelvak en deelvak geen gemeenschappelijke items die gebruikt kunnen worden om de cijfers van Economie 1,2 te transformeren naar de cijferschaal van Economie 1. Voor Scheikunde telt bovendien dat de gemiddelde correctie klein is (-0.1 in 2002, -0.2 in 2003, en 0.1 in 2004). Bij Natuurkunde is de gemiddelde correctie in 2003 en 2004 wel groot (0.7),

³² Bij een logistische regressieanalyse waarbij eerst een blok met twee dummy variabelen voor de drie opleidingen werd opgegeven, viel de constante zelfs nog iets hoger uit.

maar de correctie is 0.0 voor de diploma's uit 2002. De relatie tussen heelvak en deelvak is bij Natuurkunde dus instabiel³³.

³³ Het is mogelijk dat de relatie zich na de startfase in 2002 stabiliseert. Dat is een aandachtspunt voor de toekomst.

8. Literatuur

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Red.) *Educational measurement, second edition*. Washington, D.C.: American Council on Education.
- Commissie Toelating Numerus Fixusopleidingen (1997). *Gewogen loting gewogen*. Den Haag: Sdu.
- De Gruijter, D.N.M. (2004). *Gewogen loting bij Psychologie; onderzoek m.b.t. de toelatingsselectie bij de opleiding Psychologie in 2004/2005*. Rapport 142, ICLON, Universiteit Leiden.
- Janssen, P.J. (2004). Selectie voor het Hoger Onderwijs; Vlaanderen medio 2004. *Onderzoek van Onderwijs, 33*, 50-56.
- Lievens, F., Coetsier, P., Janssen, P.J., & Decaestecker, C. (2001). Predictieve validiteit en sekse-specificiteit van het toelatingsexamen “Arts en Tandarts” in Vlaanderen: Een eerste peiling. *Pedagogische Studiën, 78*, 4-15.
- Meerum Terwogt-Kouwenhoven, K. (1980). De voorspelbaarheid van studieprestaties van eerstejaars studenten aan de subfaculteit psychologie van Amsterdam. *Tijdschrift voor Onderwijsresearch, 5*, 126-138.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: American Council on Education, pp. 13-103.
- Onderwijsraad. (2005). *De stand van educatief Nederland*. Den Haag.
- Pijl, Y. J. (1991). *Discrepantie tussen het schoolonderzoek en het centraal schriftelijk examen in havo en vwo*. Groningen: Gion.
- Roeleveld, J. (1997). Lotingscategorieën en studiesucces. Bijlage 2 bij het rapport van de Commissie Toelating Numerus Fixusopleidingen, *Gewogen loting gewogen*. Den Haag: Sdu.

- Scherft, J. P., & Van Hilten, W. (1980). Gemiddeld eindexamencijfer en studieresultaten tot en met het kandidaatsexamen van Leidse medische studenten. *Universiteit en Hogeschool*, 26, 349-362.
- Smits, N, Mellenbergh, G.J., & Vorst, H.C.M. (2002). Alternative missing data techniques to grade point average: imputing unavailable grades. *Journal of Educational Measurement*, 39, 187-206.
- Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Van den Bergh, H., Rohde, E., & Zwarts, M. (2003). Is het ene examen het andere? Over de stabiliteit van schoolonderzoek en centraal examen. *Pedagogische Studiën*, 80, 176-191.
- Van Houwelingen, J.C., Stijn, Th., & van Strik, R. (1995). *Inleiding tot de medische statistiek*. Utrecht: Bunge.
- Van Schooten, E., & De Glopper, K. (2002). Dalende leerlingprestaties op de centraal schriftelijke examens Duits, Engels en Frans in mavo, havo en vwo? *Pedagogische Studiën*, 79, 5-17.

Bijlage 1: De cijfers bij de opleiding en het behalen van de propedeuse in 1 en 2 jaar

In het hoofdonderzoek zijn de bij de onderzochte opleidingen behaalde cijfers niet onderzocht. De opleiding Psychologie heeft ten behoeve van het onderzoek echter ook gegevens met betrekking tot de behaalde cijfers voor het onderzoek ter beschikking gesteld. Het gaat om de cijfers behaald door de studenten uit de cohort 2002.

Bij Psychologie doet in de propedeusefase niet iedereen dezelfde vakken: major-minor studenten hebben ook in het eerste studiejaar vakken bij hun minor. Daarom is besloten het deelonderzoek te beperken tot cijfers voor de vakken die in alle pakketten zitten – *Tutoraat Academische Vaardigheden, Inleiding in de M&T, Inleiding in de Psychologie, Toetsende Statistiek* en *Experimentele Proefopzettenleer* - én het vak *Geschiedenis van de Psychologie*, dat ook door praktisch iedereen gedaan wordt. In het gemeenschappelijke deel van zes vakken, ter waarde van iets meer dan de helft van het aantal ECTS in het eerste studiejaar^{B1.1}, zijn de M&T vakken oververtegenwoordigd.

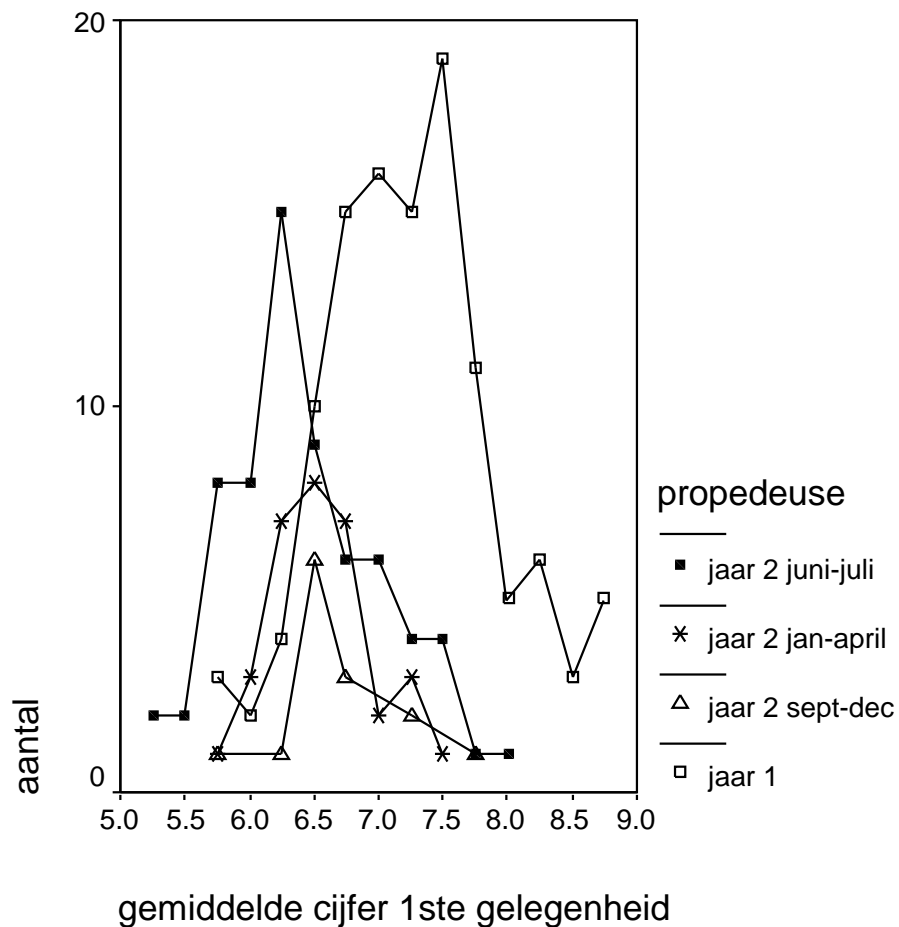
De gegevens kunnen gebruikt worden om na te gaan in hoeverre de studenten die de propedeuse op schema halen, zich ook anderszins qua niveau onderscheiden van de overige studenten. Voor de deelgroep van 226 studenten die de propedeuse binnen twee jaar hebben behaald en die op de zes genoemde vakken voor maximaal één vak geen cijfer (niet vereist in de opleidingsvariant, vrijstelling) hebben, is de relatie tussen het gemiddelde cijfer en het in het eerste of tweede jaar behalen van de propedeuse onderzocht^{B1.2}.

Van de 226 studenten waren er 114 die de propedeuse in het eerste studiejaar hebben behaald. De overige 112 studenten behaalden de propedeuse in de loop van het tweede studiejaar (het tijdstip waarop studenten in het tweede jaar kunnen slagen, hangt af van de tentamendata van de vakken waarvoor zij nog geen voldoende hebben) of aan het eind van het tweede studiejaar. Het gemiddelde cijfer

^{B1.1} De genoemde vakken worden voor alle major-minor studenten aangevuld met enkele andere Psychologievakken en vakken uit de desbetreffende minor.

^{B1.2} Van de groep studenten die de propedeuse niet in twee jaar had behaald, waren er slechts 57 met een cijfer voor minstens 5 van de 6 vakken. De gemiddelde cijfers in deze groep waren laag, op één uitzondering na. Zoals al eerder is opgemerkt, kunnen ook studenten die de studie aankunnen om hen moverende redenen met de studie stoppen.

bij de eerste tentamengelegenheid (de eerste keer dat de student tentamen deed) bedroeg 7.3 voor de studenten die de propedeuse in het eerste jaar haalden (bij een standaardafwijking van 0.67), terwijl de overige studenten een gemiddeld cijfer van 6.5 haalden (bij een standaardafwijking van 0.54). Het verschil is niet alleen op 1%-niveau significant, maar ook groot. In Figuur B1.1 is te zien hoe sterk de groep studenten die in het eerste jaar de propedeuse haalt, verschilt van de groep studenten die de propedeuse later haalt; in de figuur is de tweede groep verder uitgesplitst naar periode van het behalen van de propedeuse in het tweede jaar. Geheel onverwacht is de uitkomst niet. Degenen die de propedeuse na het eerste studiejaar halen, kunnen tentamens hebben uitgesteld tot het tweede studiejaar, maar ook vertraagd zijn omdat zij niet alle tentamens waarvoor zij een onvoldoende hebben behaald, in het eerste studiejaar op een voldoende niveau hebben kunnen brengen. Als wij de twee groepen vergelijken op basis van de gemiddelden berekend over de hoogste cijfers per vak, wordt het verschil tussen beide groepen kleiner. Het verschil verdwijnt echter niet.



Figuur B1.1. De relatie tussen gemiddeld cijfer en tijdstip propedeusediploma

Bijlage 2: Het koppelen van de examens van het deelvak en het heelvak

A.A. Béguin

In het VWO bestaan voor elk van de vakken Wiskunde A, Wiskunde B, Natuurkunde, Scheikunde en Economie twee verschillende examens, het zogeheten deelvak en heelvak. Beide examens zijn bedoeld als afsluiting van een onderwijsprogramma, waarbij het deelvak en het heelvak van elkaar verschillen in de inhoud van het programma, een verschillende studielast hebben en voorbereiden op andere vervolgopleidingen. Toch zijn er overlappen tussen de verschillende onderwijsprogramma's van deelvak en heelvak waardoor in het afsluitende examen heel- en deelvakken een aantal vragen gemeenschappelijk hebben. Uitzondering hierbij is het examen economie waarbij het deelvak geen overlap heeft met het heelvak. In deze bijlage wordt uiteengezet hoe ten behoeve van het onderzoek naar de selectie van aankomende studenten, de cijfers op het examen van het heelvak zijn omgezet in cijfers op het deelvak. Uitdrukkelijk moet opgemerkt worden dat deze omzetting een vertaling impliceert van de inhoud van beide examenprogramma's naar een gemeenschappelijke schaal. Bij de constructie van de examens en de samenstelling van de onderwijsprogramma's is er niet naar gestreefd om deze vertaling naar een gemeenschappelijke schaal mogelijk te maken. Het kan dus zijn dat beide examens minder goed op een gemeenschappelijke schaal passen. Het doel is echter te onderzoeken of vanuit examencijfers van heel en deelvak een gemeenschappelijke cijferschaal geconstrueerd kan worden die een betere voorspeller is van studiesucces dan de ongecorrigeerde cijferschalen. En voor dit doel lijkt omzetting van cijfers op basis van de overlappende vragen in ieder geval gerechtvaardigd.

Bij de omzetting van cijfers op de examens van deelvak en heelvak wordt gebruik gemaakt van de resultaten van een steekproef van kandidaten op beide examens. Zo een steekproef van kandidaten op examens wordt verzameld ten behoeve van de normering en kwaliteitscontrole van de examens (zie Alberts & Noordermeer, 2003). Bij de vakken in dit onderzoek bestaat de steekproef uit ongeveer 2000 kandidaten per examen. Van deze kandidaten worden de deelscores per vraag verzameld. In Tabel B2.1 staat de grootte van de steekproef voor de heel- en deelvakken en het aantal te behalen scorepunten in de examens en de overlap tussen de examens.

Tabel B2.1. gegevens over heel- en deelvakken bij het eerste tijdvak van het examen 2003

	Aantal kandidaten in de steekproef		Maximum te behalen scorepunten		Scorepunten in de overlap
	heelvak (1,2)	deelvak (1)	heelvak (1,2)	deelvak (1)	
Wiskunde A	2282	1986	90	84	39
Wiskunde B	2024	2056	86	83	42
Scheikunde	2066	1996	70	65	33
Natuurkunde	1932	2078	81	79	40
Economie	2214	1848	62	63	0

Met de data verzameld in de steekproeven hebben we een zogeheten ‘*verbonden onvolledig design*’ (Verhelst, 1993, p.166). De karakteristieke kenmerken hiervan zijn dat niet alle kandidaten alle vragen maken (=onvolledig) en dat er overlap bestaat tussen de vragen die de verschillende groepen kandidaten maken (=verbonden). Op deze data kunnen verschillende technieken worden toegepast om de resultaten van het ene examen te vertalen naar de resultaten op het andere examen. Welke techniek geschikt is hangt af van de eigenschappen waaraan de beide examens geacht worden te voldoen en de mate van nauwkeurigheid van de gewenste omzetting. In ons geval gebruiken we een techniek die bekend staat als ‘*calibration*’ (Mislevy, 1992). Hierbij schatten we een statistisch model op basis van de data in onze dataset, waarbij in kaart gebracht wordt hoe de moeilijkheidsgraad van de vragen van beide examens zich tot elkaar verhouden en hoe de populatie van kandidaten op het examen zich tot elkaar verhouden in prestatieniveau. Deze vorm van modelleren wordt *Item Response Theorie* (IRT) genoemd. In dit geval gebruiken we een vrij restrictief maar eenvoudig IRT-model het zogeheten ‘*partial credit model*’ (zie bijvoorbeeld Verhelst 1993). Dit model is geschikt voor onder andere ‘open vragen’ waarbij het antwoord van de kandidaat ook gedeeltelijk correct kan zijn. In het geval van een gedeeltelijk correct antwoord krijgt de kandidaat een gedeelte van de maximum score voor het volledig correcte antwoord. In het partial credit model wordt per vraag voor elke scorepunt dat behaald kan worden een parameter geschat. Deze parameter kan geïnterpreteerd worden als de moeilijkheidsgraad van het behalen van dit scorepunt (wanneer het vorige scorepunt al behaald is). In een meer flexibel statistisch model wordt als extra parameter per vraag ook een zogeheten ‘discriminatieparameter’ geschat. Deze parameter geeft aan hoe snel extra kennis of vaardigheid leidt tot een groter kans op het behalen van extra scorepunten. Het gebruik van deze extra parameter kan leiden tot een betere beschrijving van de data. Door het relatief grote aantal kandidaten in de steekproef zijn de toetsstatistieken (zie Verhelst, 1993) bij beide type modellen zeer gevoelig en registreren ze ook schendingen die het gevolg zijn

van relatief kleine afwijkingen van het model. In onze data vinden we dan ook duidelijke modelschendingen. Uit robuustheidsonderzoek (Béguin, 2000) waarbij nog flexibeler multidimensionele modellen werden gebruikt is echter naar voren gekomen dat de invloed van het gebruikte model slechts een geringe invloed heeft op de resultaten van normhandhaving van examens. In de hier gebruikte toepassing volstaat daarom ook het gebruik van het partial credit model.

Op basis van de geschatte moeilijkheidsgraad van de vragen en prestatieniveau van de populatie kandidaten op het heel- en deelvak kan vervolgens voorspeld worden hoe de populatie kandidaten van het deelvak het heelvakexamen gemaakt zou hebben (of hoe de populatie kandidaten van het heelvak het deelvak gemaakt zou hebben). Hieruit kan worden afgeleid hoe de scores op heel- en deelvak in elkaar zijn om te zetten. Hierbij maken we gebruik van een omzetting via de geschatte scoreverdeling op de examens, een techniek die ook wel ‘observed score equating of number correct scores’ genoemd wordt^{B2.1} (Lord & Wingersky, 1984).

Afhankelijk van de populatie waarvoor de scoreverdelingen op beide examens worden bepaald en afrondingen door het discrete karakter van de scoreverdeling kunnen er kleine verschillen optreden tussen de transformatie van deelvak naar heelvak en de transformatie van heelvak naar deelvak. Voor een meer uitgewerkte beschrijving van deze procedure wordt verwezen naar (Glas & Béguin, 1999 of Béguin, 2000).

Referenties

Alberts, R.V.J. & Noordermeer, L. (2003). *Examenverslag 2003*. Arnhem: Cito.

Béguin, A.A. (2000) *Robustness of Equating High-Stakes Tests*. Enschede: Universiteit Twente.

Glas, C.A.W. & Béguin, A.A. (2000). Normhandhaving bij examens in het voortgezet onderwijs: een toepassing van Item-Response theorie. In J.M. Pieters, T. Plomp & L.E. Odenthal (Red.), *20 jaar Toegepaste Onderwijskunde*: Enschede: Twente University Press.

^{B2.1} Als alternatief hiervoor geldt de zogeheten ‘true-score equating’ waarbij de gelijkheid tussen de examens wordt gedefinieerd op basis van een niet observeerbare latente vaardigheid. In het geval van ‘observed score equating’ wordt gebruik gemaakt van de eigenschappen van de scoreverdeling. (zie Béguin, 2000. p.11)

Lord, F.M. & Wingersky, M.S. (1984). Comparison of IRT true-score and equipercentile observed-score “equatings”. *Applied Psychological Measurement*, 8, 453-461.

Mislevy, R.J. (1992). *Linking educational assessments: Concepts, issues, methods and prospects*. Princeton, N.J.: ETS Policy Information Center.

Verhelst, N.D. (1993). Itemresponstheorie. In T.J.H.M. Eggen & P.F. Sanders (Red.) *Psychometrie in de praktijk*. Arnhem: Cito.

ICLON, Afdeling Hogeronderwijs

Wassenaarseweg 52
Postbus 9555
2300 RB Leiden

T 071 527 7170
F 071 527 7181
E-mail: iclonso@ICLON.LeidenUniv.nl