



Toetsing & Examinering

Juridische & Onderwijskundige
grondslagen

◆ Henk van Berkel

Inhoud

INLEIDING	12
HOOFDSTUK 1	
HET ONDERWIJSKUNDIGE EN HET JURIDISCHE DOMEIN	16
1.1 De studie van Cohen	18
1.2 Een korte geschiedenis van het examineren	24
1.3 Examinering op universiteiten	27
1.4 Juristen en onderwijskundigen	33
1.5 De verwevenheid van juristen en onderwijskundigen	36
1.6 Samenvatting	38
HOOFDSTUK 2	
DE RICHTLIJNEN: INTRODUCTIE	45
2.1 Herkomst	46
2.2 Stappen in het toetsproces	46
2.3 Samenvatting	47
HOOFDSTUK 3	
LEERDOELEN	49
3.1 Inleiding	50
3.2 Classificatie	51
3.2.1 Taxonomie	52
3.2.2 Eenvoudige taxonomie	54
3.3 Samenvatting	55
HOOFDSTUK 4	
TOETSMATRIJS = BLAUWDruk + NIVEAU	57
4.1. Blauwdruk	58
4.2 Toetsmatrijs	59
4.3. Samenvatting	61



HOOFDSTUK 5

TOETS- EN VRAAGVORMEN	62
5.1 Traditionele examen- en toetsvormen	63
5.2 Alternatieve toetsvormen	66
5.3 Open en gesloten vragen	69
5.3.1 Open versus gesloten vragen	70
5.3.2 Belangrijke onderwerpen	72
5.3.3 Inspiratiebronnen voor het construeren van toetsvragen	73
5.3.4 Aantal vragen (toetslengte) en toetstijd	75
5.3.5 Gebruikte toetsvragen	81
5.3.6 Meer vragen construeren dan nodig	85
5.3.7 Itembank	86
5.3.8 Plaatjes bij toetsvragen	87
5.4 Samenvatting	88

HOOFDSTUK 6

GESLOTEN TOETSVRAGEN	91
6.1 Juist/onjuist-vragen	93
6.2 Meerkeuzevragen	94
6.2.1 Alternatieven: aantal en volgorde	96
6.2.2 Construeren van gesloten toetsvragen	96
6.3 Checklist gesloten vragen	112
6.4 Samenvatting	113

HOOFDSTUK 7

OPEN TOETSVRAGEN	116
7.1 Werkwoordvormen	117
7.2 Typen open vragen	118
7.3 Sturen van de lengte van het antwoord	119
7.4 Construeren van open vragen	121
7.5 Checklist open vragen	125
7.6 Samenvatting	126

HOOFDSTUK 8

ANTWOORDMODELLEN	128
8.1 Het belang van een antwoordmodel	129
8.2 De vorm en inhoud van een antwoordmodel	131
8.3 Taalfouten	132
8.4 Voorbeeld van een antwoordmodel	133
8.5 Samenvatting	137



HOOFDSTUK 9

CONTROLEREN	139
9.1 Inleiding	140
9.2 Empirische controle (try-out)	141
9.3 Kwaliteitsbeoordeling vooraf	142
9.3.1 Beoordelaars	142
9.3.2 Beoordelingsaspecten	143
9.3.3 Beoordelingsmethoden	144
9.4 Samenvatting	146

HOOFDSTUK 10

SAMENSTELLEN VAN DE TOETS	148
10.1 Toetsmatrijs	149
10.2 Samenvatting	150

HOOFDSTUK 11

DE AFNAME VAN DE TOETS	152
11.1 Toetsinstructie	155
11.2 Toetstijd	156
11.3 Invulinstructie	156
11.4 Scoring en uitslag	157
11.4.1 Punten per vraag	158
11.4.2 Cijferbepaling	158
11.4.3 Uitslagtermijn	160
11.5 Algemene regelingen	160
11.5.1 Hulpmiddelen	160
11.5.2 Rekenmachientjes	162
11.5.3 Nummering bladzijden	162
11.5.4 Vermelding naam	163
11.5.5 Frauduleus handelen	163
11.5.6 Vragen tijdens de toets	163
11.5.7 Laatkomers en vroege vertrekkers	164
11.5.8 Toiletbezoek	165
11.5.9 Toetsevaluatie	166
11.5.10 Inleveren	168
11.5.11 Voorbeeld van een toetsinstructie	168
11.6 Lay-out van de toets	171
11.7 Toetsorganisatie	172
11.7.1 Toetsruimte	172



11.7.2	Verstoringsen	173
11.7.3	Verstrekking van koffie, thee, et cetera	174
11.7.4	Verantwoordelijkheid tijdens de toets	175
11.7.5	Surveillanten	176
11.8	Maatregelen tegen fraude	177
11.8.1	Controle identiteit	177
11.8.2	Vraagvolgorde	177
11.8.3	Alert surveilleren	178
11.8.4	Fraudereglement	178
11.8.5	Tassen en jassen	178
11.8.6	Oordopjes, horloges, koptelefoons, googlebrillen, et cetera	179
11.8.7	Verplichte plaatsen	179
11.8.8	Algemene maatregelen	179
11.9	Samenvatting	182

HOOFDSTUK 12

BEOORDELEN

185

12.1	Scoren en beoordelen van gesloten vragen	186
12.1.1	Machinaal beoordelen	186
12.1.2	Raadkans	187
12.1.3	Gewichten toekennen aan vragen	188
12.2	Scoren en beoordelen van open vragen	189
12.2.1	Wat er mis kan gaan	189
12.2.2	Oplossingen	191
12.3	Verschillen in beoordelingen	198
12.4	Afronden	198
12.5	Globaal versus analytisch beoordelen	201
12.6.	Rubrics	203
12.6.1	Structuur van een rubric	204
12.6.2	De voor- en nadelen	204
12.7	Beoordelingen op schrift en transparant	205
12.8	Samenvatting	206

HOOFDSTUK 13

ANALYSEREN

210

13.1	Bij een combinatie-toets: twee analyses	212
13.2	Wie verricht de psychometrische analyse?	212
13.2.1	Hulp van een computer	212
13.2.2	Toetsservicebureau	213



13.3	Invoer van data	214
13.4	De -1, 0, 1 scoring van meerkeuzevragen	216
13.5	Wat moet worden berekend?	216
13.6	Begrippen	217
13.6.1	Deelscores en eindscore	217
13.6.2	Vraagmoeilijkheid (p-waarde) en toetsmoeilijkheid	218
13.6.3	Afleidings van meerkeuzevragen (a-waarde)	223
13.6.4	Onderscheidingsvermogen van toetsvragen (<i>Rit</i> , <i>Rat</i> , <i>Rar</i> en <i>Rir</i>)	225
13.7	De betrouwbare van een psychometrische analyse	234
13.8	Toetsbetrouwbaarheid	234
13.8.1	Benaderingen voor de test-hertest betrouwbaarheid	235
13.8.2	Hoe hoog moet de toetsbetrouwbaarheid zijn?	239
13.8.3	Bepaling van de toetslengte (Spearman-Brown formule)	243
13.8.4	Reparatiemogelijkheden	245
13.9	Validiteit	248
13.10	Relatie tussen betrouwbaarheid en validiteit	250
13.11	Samenvatting	250

HOOFDSTUK 14

NORMEREN EN HET GEVEN VAN CIJFERS		254
14.1	Normeren en het geven van cijfers is problematisch	255
14.2	Het kiezen van een beoordelingsschaal	258
14.3	Normstelling bij het geven van cijfers	261
14.3.1	Relatief normeren	261
14.3.2	Absoluut normeren	264
14.3.3	Absoluut en relatief normeren vergeleken	265
14.4	Bepaling van de zak-/slaaggrens	266
14.4.1	Opvattingen over wenselijke zak-/slaagpercentages	266
14.4.2	Methoden voor het bepalen van de zak-/slaaggrens	269
14.5	Enkele algemene eisen aan het normeren	283
14.6	Samenvatting	286

HOOFDSTUK 15

RAPPORTEREN		289
15.1	Rapportage aan studenten	291
15.1.1	Autonomie en feedback	291
15.1.2	Tijdstip van feedback	292



15.1.3	Inhoudelijke feedback	294
15.2	Rapportage aan de opleiding	294
15.3	Samenvatting	297

HOOFDSTUK 16

INZAGE EN NABESPREKING		299
16.1	Inzage en nabespreking	300
16.2	Wat zijn de voor- en nadelen?	301
16.3	De organisatie	303
16.4	Hoe te voorkomen dat studenten er met de vragen vandoor gaan?	307
16.5	Samenvatting	307

HOOFDSTUK 17

TOETSVRAGEN CONSTRUEREN MET ARTIFICIËLE INTELLIGENTIE?

<i>Vraag het ChatGPT, of toch maar niet?</i>		310
17.1	De door ChatGPT geconstrueerde toetsvragen	312
17.1.1	Meerkeuzevragen: Feitenkennis	312
17.1.2	Meerkeuzevragen: Toepassing	313
17.1.3	Meerkeuzevragen: Inzicht	315
17.2.1	Kort-open toetsvragen: Feitenkennis	316
17.2.2	Kort-open toetsvragen: Toepassing	317
17.2.3	Kort-open toetsvragen: Inzicht	319
17.2.4	Reflectievragen	320
17.3	Conclusie	321
17.4	Samenvatting	322

HOOFDSTUK 18

SAMENVATTING		324
Referenties		328
Uitspraken van Beroepscolleges		336
Index		337
Bijlage 1: Begrippen		242
Bijlage 2: Curriculum Vitae van Henk van Berkel		358



Inleiding

In het maatschappelijk leven nemen examens een belangrijke plaats in. Het zijn proeven van bekwaamheid met veelal grote gevolgen. Slagen betekent dat studenten hebben bewezen te voldoen aan bepaalde eisen. Dat geeft hun niet alleen een status, maar voldoende gemaakte toetsen of examens geven bijvoorbeeld ook toelating tot een vervolgopleiding, toelating tot een bepaald beroep, tot het mogen blijven uitoefenen van een bepaald beroep, tot het mogen rijden van een auto, et cetera. Maar ook voor de maatschappij zijn examens belangrijk. Slagen voor een examen is een bewijs dat afgestudeerden zekere competenties hebben verworven die zij ten dienste kunnen stellen van de maatschappij. De maatschappij moet daarom vertrouwen kunnen hebben in de examens. Dus zowel examinandi als de maatschappij hebben belang bij examens. De kwaliteit van examinering is daarom van eminent belang.

Er zijn weinig activiteiten in het dagelijkse leven waarin niet wordt gemeten. Sommige metingen zijn direct. Dat wil zeggen, de meetlat wordt naast een voorwerp gelegd en de maat wordt eraan afgelezen, de lengtemeting van personen bijvoorbeeld. Andere metingen zijn indirect, zoals het aflezen van de temperatuur op een thermometer. De thermometer reageert op een bepaalde manier op temperatuur en maakt daardoor de temperatuur zichtbaar.

Examens kunnen directe metingen zijn, maar ook indirecte. Als ze een beroep doen op de kennis van studenten, zijn het indirecte metingen. Het is niet mogelijk kennis van personen af te zonderen en uit te spreiden om er vervolgens een meetlat langs te leggen. Kennis is iets abstracts. Het is een latente trek, niet direct zichtbaar. Toch wil een examiner weten hoeveel kennis een student heeft over een bepaald onderwerp. Met andere woorden, er zou een soort instrument moeten zijn dat de kennis representeert. Een dergelijk instrument heeft talloze benamingen: proefwerk, overhoring, tentamen, et cetera. Een verzamelaanduiding voor al deze benamingen is toets. Toetsen zijn



instrumenten om kennis te meten. Het meten van vaardigheden maakt vaak een directere meting mogelijk. Schrijfvaardigheden bijvoorbeeld blijken uit een schriftelijk stuk dat is geproduceerd. Dat laat zich beoordelen en is daardoor een directe meting.

Examens

Er wordt wel eens beweerd dat in Nederland het ene deel van de bevolking het andere examineert. Of dat waar is, is nooit onderzocht, maar het zal zeker enigszins overdreven zijn. Dat laat onverlet dat er heel wat wordt geëxamineerd. Vooral wanneer ook de (tussentijdse) tentamens en schooltoetsen in het onderwijs meetellen. Alom bekende categorieën zijn:

- In het voortgezet onderwijs (vmbo, havo en vwo) doen jaarlijks 200.000 leerlingen het centraal schriftelijk eindexamen; daarnaast maken zij tijdens het onderwijs nog tal van toetsen, proefwerken bijvoorbeeld.
- In het middelbaar beroepsonderwijs (mbo) leggen jaarlijks tienduizenden het eindexamen af; ook zij maken tussentijdse toetsen op school.
- Honderdduizenden studenten maken binnen een studiejaar diverse tentamens.
- Het Centraal Bureau Rijvaardigheidsbewijzen neemt jaarlijks meer dan een miljoen theorie- en praktijkexamens af.
- Tienduizenden personen doen taaltoetsen om hun taalvaardigheid te bewijzen.
- Inburgeringsexamens zijn verplicht voor nieuwkomers in Nederland en jaarlijks zijn er duizenden kandidaten.

Daarnaast zijn er uiteraard de kwalificerende examens om een beroep te mogen uitoefenen. Die examens zijn uiterst divers. Denk hierbij aan makelaars, financieel deskundigen, banketbakkers, opticiens, et cetera. Veruit de meeste beroepen vereisen een (starters)diploma. Maar een kwalificerend examen is doorgaans geen eenmalige exercitie. De meeste beroepsgroepen eisen dat personen hun beroep onderhouden en daarvan een bewijs leveren; zij moeten dus weer een examen afleggen.



Het zal niet overdreven zijn te beweren dat er jaarlijks bij miljoenen personen examens worden afgenomen. Lang niet iedere persoon krijgt een individueel examen. Hetzelfde examen wordt bij meerdere personen en vaak gelijktijdig afgenomen. Al die examens moeten worden geconstrueerd. Daarover gaat dit boek. Het boek volgt in virtuele zin een examen- of toetsconstructeur bij het maken van leerdoelen tot en met het evalueren van het afgelegde examenproces. Dat is een lange weg die zorgvuldigheid vereist.

Summatief en formatief

De meeste opleidingen hanteren toetsen om na te gaan of studenten voldoende weten of kunnen van het onderwerp waar de toets over gaat. Deze functie van toetsen wordt aangeduid als de summatieve functie. Een summatieve toets sluit als het ware iets af. Het is een proeve van bekwaamheid. De uitslag dient het niveau te weerspiegelen van de student. De hoogte van het niveau bepaalt de beslissing. De meest gangbare beslissingen kennen benamingen als: slagen/zakken, overgaan/blijven zitten, voldoende/onvoldoende, toelaten/weigeren, et cetera.

Een term die vaak in één adem wordt genoemd met summatief is formatief. Formatief toetsen, een betere term is formatief evalueren, is het doelgericht ontlokken van reacties over het leren en leerproces van studenten. Deze reacties worden vervolgens geïnterpreteerd én gebruikt om betere beslissingen te nemen over vervolgstappen in leren en/of instructie. Het gaat er bij formatief evalueren in de kern om dat de informatie die wordt verzameld over het leren en het leerproces, wordt gebruikt om goed onderbouwde beslissingen te kunnen nemen over vervolgstappen in het onderwijs en het leren van de studenten. Aan formatief evalueren zijn geen consequenties verbonden in examenteknische zin. Het beïnvloedt niet direct de studievoortgang. De studenten kunnen aan de hand van de formatieve uitslag bepalen waar hun hiaten in de leerstof zitten. Daarvoor is uiteraard nodig dat de uitslag voldoende specifiek en gedetailleerd is. Ook voor examinatoren is formatieve evaluatie relevant. Uitslagen geven aan in welke mate het onderwijs een opbrengst heeft gehad. Welke onderwerpen zijn goed blijven hangen in de hoofden van studenten? Van welke onderwerpen weten zij weinig af, terwijl dat wel zou moeten? Waar moet het volgend jaar meer op worden gehamerd?



Waar het om gaat is dat de test-/examentheorie zich in de loop der jaren heeft ontwikkeld tot een erkende discipline die zich stevig heeft genesteld in de maatschappij. Het werk van Binet & Simon is opgepakt door vele vooraanstaande onderzoekers. Zij hebben de intelligentietest als startpunt genomen voor het bouwen van een theorie. Het bleek al gauw dat dezelfde theorie ook als basis kon dienen voor het ontwikkelen van kwalificerende en selecterende examens, en, daaruit afgeleid, andere vormen van toetsen. Voeg daaraan toe het werk van auteurs die de theorie wisten om te zetten in handreikingen voor degenen die in de praktijk examens gingen ontwikkelen, en er ontstaan enerzijds theoretische onderbouwingen en anderzijds praktische toepassingsmogelijkheden. Het geheel vormt een doortimmerd bouwwerk. Dit bouwwerk vormt de basis voor dit boek. Dat geldt in ieder geval voor het onderwijskundig domein. In hoofdstuk 1 wordt uitgelegd wanneer het tweede domein, het juridische, is ontwikkeld en hoe en wanneer de twee domeinen min of meer zijn geïntegreerd.





HOOFDSTUK 1

HET ONDERWIJSKUNDIGE EN HET JURIDISCHE DOMEIN

- » *Wat is de bijdrage van het proefschrift van Cohen uit 1981 aan de toets- en examenpraktijk in het hoger onderwijs?*
- » *Welke drie invalshoeken bespreekt Cohen met betrekking tot de studierechten van studenten in het wetenschappelijk onderwijs?*
- » *Wat is het zorgvuldigheidsbeginsel en hoe wordt het toegepast in het kader van examineren in het hoger onderwijs?*
- » *Welke rol spelen de Colleges van Beroep en de Raad van State in het toets- en examenproces?*
- » *Wat is het verschil tussen het formele en materiële beginsel in het bestuursrecht zoals toegepast op examens?*
- » *Wat is het belang van het gelijkheidsbeginsel bij het beoordelen van toetsen en examens?*
- » *Hoe beïnvloeden juristen en onderwijskundigen gezamenlijk de examenreglementen en de examenpraktijk aan universiteiten?*
- » *Wat zijn de kenmerken van een selecterend examen en hoe verschilt dit van een kwalificerend examen?*
- » *Welke impact heeft de toegenomen juridisering op de academische vrijheid van examinatoren in het hoger onderwijs?*

Dit boek bevat richtlijnen voor de toets- en examenpraktijk. De richtlijnen zijn afkomstig uit twee domeinen: het onderwijskundige en het juridische. De bestudering van de toetspraktijk vanuit het juridische domein is begonnen met het proefschrift van Cohen uit 1981¹ en het onderwijskundige, meer in het bijzonder het toetstechnische domein, komt voort uit theorieën over psychologische testen die zich ruim honderd jaar geleden begonnen te ontwikkelen. Dit boek bestrijkt beide domeinen. Het zal uit beide domeinen richtlijnen presenteren die bruikbaar zijn bij het inrichten van examen- en toetsprocessen. Na de studie van Cohen zijn beroepscolleges in het leven geroepen die vanuit de juridische invalshoek geschillen hebben beoordeeld: Het College van Beroep voor de Examens (CBE) aan iedere instelling van het hoger onderwijs en het College van Beroep voor het Hoger Onderwijs (CBHO), later, in 2023, ondergebracht bij de Raad van State (RvS). Uit de uitspraken over de geschillen door de beide beroepscolleges zijn richtlijnen af te leiden. Deze worden gepresenteerd en besproken in hoofdstukken 2 t/m 16.

¹ Cohen 1981.



1.1 De studie van Cohen

Het proefschrift van Cohen ligt op het terrein van twee domeinen: het recht en de onderwijskunde. Cohen verrichtte een belangwekkende studie naar de rechtspositie en rechtsbescherming van studenten in het hoger onderwijs, in het bijzonder het wetenschappelijk onderwijs (wo). Het is een baanbrekend werk omdat het de eerste studie is die een verband probeert te leggen tussen enerzijds de onderwijskundige maatstaven ten aanzien van examinering en anderzijds de maatstaven voortkomend uit het recht.

In zijn proefschrift bespreekt Cohen langs drie invalshoeken de studierechten van studenten in het wetenschappelijk onderwijs: 1. De rechten van studenten bij de toelating tot de universiteit; 2. De rechten van studenten ten aanzien van het verkrijgen van onderwijs; 3. De rechtspositie van studenten bij examens. De derde invalshoek is voor dit boek relevant. Cohen betoogt allereerst dat toetsuitslagen beschikkingen zijn van een bestuursorgaan. Ze vallen daarom onder het bestuursrecht en de regels die daaruit afkomstig zijn: de (on)geschreven regels en rechtsbeginselen. Geschreven rechtsregels over het beoordelen zijn er, vooral ten tijde van zijn proefschrift, nauwelijks, dus stapt Cohen al snel over naar de ongeschreven rechtsbeginselen. Bieden die aanknopingspunten voor het formuleren van regels waaraan toetsen dienen te voldoen?

Zijn onderzoeksmethode is dat hij de toetstheoretische eisen (voornamelijk geformuleerd door de vooraanstaande Nederlandse toetsdeskundige A.D. de Groot) als norm neemt waaraan hij formele rechtsregels koppelt, en ook de algemene beginselen van behoorlijk bestuur, om vervolgens tot materiële rechtsregels te komen. Hij onderzoekt in hoeverre de regels corresponderen met de bestaande werkelijkheid zoals deze tot uiting komt in de uitspraken van de beroepscommissies die vergelijkbaar zijn met de huidige CBE's. Dat leidt ertoe dat de formuleringen van een aantal regels worden gewijzigd. Met andere woorden, Cohen geeft materiële inhoud aan de formele regelingen zoals die in die tijd bestonden. De zo ontwikkelde rechtsregels kunnen als toetssteen worden gebruikt bij het vaststellen van de rechten van studenten.



Zijn studie heeft daarmee betrekking op de materiële studierechten. Cohen heeft zich laten leiden door onderwijskundige argumenten, door rechtsbeginselen en door jurisprudentie, voor zover aanwezig. Naast dat hij die uitspraken hanteert als input voor het formuleren van rechtsregels, stelt hij zich ook kritisch op ten aanzien van de uitspraken van beroepscolleges. Sommige daarvan vindt hij apert onjuist, wat wil zeggen niet kloppend met de (interpretatie van) geldende regelingen. In zekere zin evalueert hij de kwaliteit van beroepszaken aan het eind van de jaren zeventig van de vorige eeuw.

In de periode tussen het voltooiën van het proefschrift van Cohen (1981) en nu, zitten meer dan veertig jaren. In die periode is het rechtssysteem met betrekking tot de rechtsbescherming van studenten verder ontwikkeld tot aan het systeem dat we nu kennen en beschreven staat in de Wet op het hoger onderwijs en wetenschappelijk onderzoek (WHW). Er is bovendien jurisprudentie ontstaan doordat CBE's, het CBHO en de RvS vele uitspraken hebben gedaan die de rechtsregels over examineren nader preciseren. Uit die preciseringen zijn richtlijnen voor examinering af te leiden.

Een tweede relevant onderwerp uit de studie van Cohen is zijn bespreking van een aantal belangrijke juridische begrippen die hij onderzoekt op toepasbaarheid in een onderwijssituatie. De algemene noemer waaronder hij deze begrippen plaatst is 'redelijkheid'. Op de eerste plaats stelt hij dat beroepscolleges het antwoord op rechtsvragen zullen proberen te vinden in schriftelijk vastgelegde regelingen. Maar die geven niet altijd een bevredigend antwoord. Vandaar dat colleges in aansluiting hierop ook beoordelen of een beslissing in strijd met de redelijkheid is. De begrippen zijn afkomstig uit het bestuursrecht en vallen onder de term 'algemene beginselen van behoorlijk bestuur'. Het zijn gedragsregels voor (overheids)organen die hun oorsprong vinden in de jurisprudentie en die deels zijn gecodificeerd. Er zijn formele en materiële beginselen. De laatste categorie heeft betrekking op de inhoud van de beslissingen. De formele beginselen moeten zijn terug te voeren op de formeel toebedeelde bevoegdheid aan het orgaan.

Hoe zijn de beginselen toepasbaar op examineren?



Zorgvuldigheid

Het zorgvuldigheidsbeginsel – Opdebeek & Van Damme noemen dit beginsel “de moeder van alle beginselen”² – speelt binnen de context van examineren in het hoger onderwijs vooral een rol in de fase voorafgaand aan de toets. Opleidingen dan wel examinatoren hebben de plicht om met zorg en nauwgezetheid de regelingen over de examinering te treffen. Dat geldt ten aanzien van de eisen, de inhoud en de vorm van de toets en het beoordelen. Cohen vindt dit beginsel weinig specifiek. In algemene zin is het een belangrijk beginsel, volgens hem, maar het zal bij de toepassing betrekkelijk weinig als grondslag kunnen dienen voor regels.

De Algemene wet bestuursrecht (Awb) verwoordt het zorgvuldigheidsbeginsel aldus:

Bij de voorbereiding van een besluit vergaart het bestuursorgaan de nodige kennis omtrent de relevante feiten en de af te wegen belangen.
(Artikel 3:2 Awb)

Dit artikel betekent dat de examencommissie bij het ontwerpen van een regeling rekening dient te houden met de gevolgen van beslissingen. Het is een formeel beginsel, net als het volgende te bespreken beginsel.

Het motiveringsbeginsel

Bij het motiveringsbeginsel binnen de context van het onderwijs staan drie vragen centraal³: 1. In hoeverre moeten beslissingen worden gemotiveerd? 2. Is de motivering voldoende om de beslissingen te dragen? 3. Is de motivering gegrond met de feiten?

Het leidt geen twijfel dat de algemene opvatting is dat beslissingen in het onderwijs die studenten raken, moeten worden gemotiveerd. Dat is niet altijd makkelijk en zelfs niet altijd mogelijk. Het voorbeeld haalt Cohen bij De Groot & Wijnen⁴. Het heeft betrekking op het bepalen van de grens tussen voldoende en onvoldoende. Het is het niet mogelijk een zak-/slaaggrens objectief te bepalen

² Opdebeek & Van Damme 2006.

³ Cohen 1981, p.87-89.

⁴ De Groot & Wijnen 1971.



(zie hoofdstuk 14). Op enig moment in het proces van cijfers geven zal de examinator het autoriteitsargument moeten gaan hanteren: "Zo heb ik het naar eer en geweten gedaan." Hij had er ook nog aan kunnen toevoegen dat zijn werkgever hem heeft aangesteld en op die wijze hem de autoriteit heeft gegeven, aldus Cohen.

Ondanks het geschetste probleem over de zak-/slaagbepaling is het motiveringsbeginsel al ver doorgevoerd in het onderwijs. Studenten hebben inzagerecht, maar ook recht op een nabespreking. De verantwoordelijke examinator moet dan uitleg geven over normen die zijn gehanteerd bij het beoordelen van het ingeleverde werk.⁵

Ook het motiveringsbeginsel heeft een plaats gevonden in de Awb en wel in artikelen 3:46 t/m 3:48.

Rechtszekerheid en vertrouwen

In navolging van Wiarda⁶ heeft Cohen de opvatting dat beide beginselen samen moeten worden genomen. Wiarda spreekt uit dat het vanzelfsprekend dient te zijn dat een orgaan zich aan de eigen regels houdt (dat geeft rechtszekerheid), net zoals het vanzelfsprekend moet zijn dat gerechtvaardigde verwachtingen worden gerespecteerd (dat geeft vertrouwen). Cohen stelt:

"Nu eens moet in verband met het vertrouwensbeginsel de rechtszekerheid gehonoreerd worden, dan weer dient die juist niet gerespecteerd te worden. Het vertrouwensbeginsel is derhalve het uitgangspunt, het beginsel der rechtszekerheid maakt daarvan deel uit, maar omvat het niet."
(p. 85)

Het rechtszekerheidsbeginsel schrijft voor dat studenten weten waar ze aan toe zijn. De WHW draagt hieraan bij door bijvoorbeeld voor te schrijven dat de opleiding een onderwijs- en examenregeling (OER) hanteert waarin talloze onderwerpen uit het onderwijs en de examinering vast komen te liggen, in

⁵ WHW, art. 7.13, lid 2, onder p en q.

⁶ Wiarda 1986.



ieder geval voor een heel studiejaar. Cohen stelt dat niet alles is vast te leggen. Vooral in de latere fase van de studie kunnen studenten zelf vormgeven aan het onderwijs, zelfs aan de inhoud. Denk daarbij bijvoorbeeld aan het onderwerp van de scriptie. Dan gaat het vertrouwensbeginsel gelden: studenten moeten erop kunnen vertrouwen dat wanneer ze werken volgens de (vage) normen van de examinerator, hun scriptie een voldoende krijgt.

Détournement de pouvoir

De term 'détournement de pouvoir' duidt op de situatie waarin een orgaan een bevoegdheid gebruikt voor een ander doel dan waarvoor zij is bedoeld. Cohen⁷ geeft het voorbeeld van een geneeskundefaculteit die het propedeutisch examen hanteerde om studenten te selecteren voor het tweede jaar waar minder plaatsen beschikbaar waren. Dat is wat anders dan een selecterende propedeuse omdat ook de studenten die de propedeuse met succes hadden voltooid – en dus bewezen kwalitatief aan de maat te zijn – afgewezen konden worden voor het vervolg van de studie. Een ander voorbeeld is de situatie dat het slagen voor de toets ook afhankelijk is van de vraag of de student de colleges heeft bijgewoond, zonder dat dat noodzakelijk is voor het bereiken van de leerdoelen. Een laatste voorbeeld van Cohen is interessant. Het is de situatie waar een student examenpunten kan verdienen door aanwezig te zijn in het onderwijs. De gegeven motivering is dan dat het volgen van onderwijs leereffecten teweegbrengt en daarom mag worden beloond. Cohen stelt vervolgens dat dat dubbelop is. Als er een leereffect optreedt van het volgen van onderwijs, dan zal dat effect ook moeten zijn terug te vinden in de toetsscore en hoeft de aanwezigheid niet afzonderlijk te worden beloond.

Willekeur

Als een orgaan een beslissing neemt die bij afweging van de relevante belangen in redelijkheid niet genomen had mogen worden, is er sprake van willekeur. Willekeur is vooral aan de orde als het orgaan dat de beslissing neemt, een grote mate van vrijheid heeft. In het hoger onderwijs hebben examineratoren een grote mate van vrijheid mits zij bij hun beoordeling binnen de leerdoelen blijven en binnen de bestaande regels. Zij beoordelen of de student voldoet

⁷ Cohen 1981, p.78-79.



aan de leerdoelen. Ook geldt dat naarmate de leerdoelen minder duidelijk zijn omschreven, de beoordelingsruimte van de examinerator groter is. In dat geval is onredelijkheid niet snel aan te tonen. Om die reden, zo gaat Cohen⁸ verder, zal dit beginsel binnen de context van beroepscolleges in het hoger onderwijs, geen grote rol spelen omdat deze colleges terughoudend toetsen. Zij moeten dus de beoordelingsvrijheid van de examinerator respecteren.

Het gelijkheidsbeginsel

Het gelijkheidsprincipe vereist dat gelijke gevallen ook gelijk worden behandeld én dat verschillende gevallen naar de mate van het verschil worden behandeld.⁹ De toevoeging is belangrijk. Het populaire gezegde ‘gelijke monniken, gelijke kappen’, dekt dus maar de helft van dit principe. Beroepscolleges zijn terughoudend om te besluiten dat er sprake is van een gelijke situatie. Een student die wijst naar een ander ‘hij wel, dus ik ook’ krijgt lang niet altijd gelijk. Toch is het niet moeilijk dit beginsel toe te passen bij het toetsen. Een schriftelijke toetssituatie is voor iedere toetsdeelnemer hetzelfde: zij krijgen meestal dezelfde vragen onder dezelfde omstandigheden. Verschillen in prestaties op de toets kunnen rechtstreeks worden toegeschreven aan verschillen in cognitieve vaardigheden, dat zich uit in verschillende waarderingscijfers, aldus Cohen.¹⁰ Verschillende toetsvormen staan correct toepassen van dit beginsel in de weg. Bij meerkeuzevragen zal duidelijk zijn dat een student die 85 meerkeuzevragen correct heeft beantwoord een hoger cijfer verdient dan een student die op dezelfde toets 63 vragen correct heeft beantwoord. Maar wordt een vak afgesloten met een mondelinge toets, dan kan een student die een hoger cijfer claimt moeilijk wijzen naar een andere student. Ook bij het beoordelen van papers speelt dit probleem.

Cohen¹¹ schetst een casus die betrekking heeft op het gelijkheidsbeginsel. Wanneer een cohort eerstejaars een toets maakt waarvoor praktisch iedere student zakt, terwijl hetzelfde cohort ook andere toetsen heeft gemaakt met ‘normale’ zakpercentages, en de groep klaagt bij de examencommissie dat de toets te moeilijk is geweest, zal de commissie moeten aantonen of aannemelijk moeten maken dat de verschillen in slaagpercentages zijn te wijten aan de

⁸ Cohen 1981, p.80-81.

⁹ Wiarda 1986.

¹⁰ In hoofdstuk 13 zal deze stellige uitspraak enigszins worden genuanceerd wanneer het begrip betrouwbaarheid aan de orde is.

¹¹ Cohen 1981, p.83-84.



verschillen in de noodzakelijk geachte cognitieve vaardigheden van de studenten. Dat is geen makkelijke klus.

1.2 Een korte geschiedenis van het examineren

In zijn oratie schetst Van der Linden¹² de herkomst van het huidige examineren. Dat gaat terug naar het China van rond het begin van de jaartelling. In dat land moesten ambtenaren worden geselecteerd. Wie mocht overheidsdienaars worden? In tegenstelling tot wat gebruikelijk was in die tijd, kwam iedere mannelijke burger daarvoor in aanmerking en niet de familieleden van de reeds zittende ambtenaren. Er moesten examens komen die persoonlijke capaciteiten in beeld konden brengen. Het waren schriftelijke examens die centraal werden afgenomen. Het vergelijkend examen bestond uit de onderdelen rechtskennis, aardrijkskunde en kennis over de landbouw. Het examen duurde enkele dagen. Deze examens functioneerden in China in min of meer gelijke vorm tot ongeveer 1900.¹³

Over een andersoortig selecterend examen is te lezen in een van de boeken van het Oude Testament. Het gaat dan over de krijgsheer Gideon die voor de taak stond de Israëlieten van de Midjanieten te bevrijden. Hij had daar uiteraard een leger voor nodig. Hoewel hem een groot leger ter beschikking stond, gaf hij er de voorkeur aan om met een kleinere groep aan te vallen. Maar welke selectiecriteria zijn hanteerbaar? Zijn methode was opmerkelijk. Hij liet de mannen water drinken uit de rivier. Degenen die met de hand het water opschepten en dronken, waren geschikt voor zijn kleinere leger (en vormden de 'Gideonsbende'). De geschiedenis vertelt jammer genoeg niet hoe Gideon ertoe kwam deze wijze van selecteren te hanteren. De 'test' heeft op het oog weinig te maken met krijgskunst.

Interessant is te constateren dat in de oudheid al twee manieren van selecteren werden toegepast: vergelijkend en absoluut. De Chinese methode selecteerde uit een groep de besten, gegeven een beperkt aantal ambtelijke functies, en Gideon koos zijn krijgers uit degenen die de taak op een, naar zijn oordeel, correcte

¹² Van der Linden 1983.

¹³ Er werden in China nog veel eerder tests afgenomen. Rond 2200 voor onze jaartelling nam de keizer zijn personeel een test af om te beoordelen of zij nog bekwaam waren. Echter van deze tests is verder niets bekend en ze worden daarom niet gezien als de eerste examens (Gregory, 2014).



manier volbrachten, zonder limiet aan het aantal uitverkorenen. De overeenkomst is dat beide examens werden gehanteerd in een selectiesituatie. Ook vandaag de dag zijn dat de twee belangrijkste vormen van selecterende examens. Denk aan rijexamens die niet vergelijkenderwijs worden afgenomen maar absoluut en aan ingangsexamens voor de studie geneeskunde op universiteiten die een gelimiteerd aantal kandidaten selecteert omdat het aantal plaatsen in de opleiding beperkt is.

Veel later, aan het begin van de twintigste eeuw, ontwikkelden Binet en zijn medewerker Simon ook een selecterend examen. Hun opdracht was een test te ontwerpen die minderbegaafde leerlingen uit Parijse scholen kon selecteren uit de hele groep leerlingen; zij kregen dan een wat eenvoudiger curriculum. De test moest objectief zijn, in ieder geval objectiever dan het oordeel van de leraren over leerlingen want dat oordeel werd verondersteld te zijn gekleurd door hun eigen opvattingen. Het selectie-instrument dat Binet & Simon ontwikkelden, werd de eerste intelligentietest.

Het zijn deze drie selectiesituaties die aanleiding hebben gegeven tot het ontwikkelen van testen die later standaardtesten worden genoemd. Dit is een type test dat ontworpen is om consistent en objectief te worden afgenomen en beoordeeld. Dit betekent dat de test onder dezelfde omstandigheden wordt uitgevoerd voor alle deelnemers, met dezelfde instructies, tijdslimieten, en vragen of taken. De beoordeling gebeurt op basis van vastgestelde criteria, zodat de resultaten op een betrouwbare manier kunnen worden vergeleken tussen verschillende individuen of groepen.

Vooraf het werk van Binet & Simon wordt gezien als de voorloper van het examineren en toetsen zoals nu plaatsvindt. Zowel examens die iets afronden (bijvoorbeeld een opleiding), als toelatingsexamens en toetsen gedurende een opleiding hebben ervan geprofiteerd. Binet & Simon konden, gezien hun opdracht om minder presterende leerlingen te midden van een grotere groep op te sporen, niet terugvallen op de gangbare psychologische onderzoeken van hun tijdgenoten als Galton, Ebbinghaus, Pearson en Spearman. Deze onderzoekers lieten hun

